


```

APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hotte, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
EARLIER FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-79

```

```

alignment_scores:
  Quality: 141.00      Length: 577
  Ratio: 0.613        Gaps: 29
  Percent Similarity: 39.861      Percent Identity: 24.957

```

```

alignment_block:
US-09-303-518D-125 x US-09-413-814-79

```

```

Align seg 1/1 to: US-09-413-814-79 from: 1 to: 1213

```

```

21 TCTAACCCTGCCATCGCGG.....CAGAC 46
   :::::|||||:::
15 AlaArgProAlaArgArgGlyAspArgArgGlyLeuAlaLeuArgSe 31
   :::::|||||:::
47 CGGACCAACCGCTTACGACGGCCGCAT..... 77
   :::::|||||:::
31 AspArgGlyLeuArgArgArgGlyHisGlyLeuGlyLeuGlyAlaG 48
   :::::|||||:::
78 TACCAAGTCGCGCTGTCGCGAGAAATATGCGGTATGCGGCC 122
   :::::|||||:::
48 LysGlyArgArgAlaArgArgAlaAlaArgArgArgLeuProAlaPro 64
   :::::|||||:::
122 .....
65 GlyProSerProArgArgGlyAlaProGlnAlaProProGlyAlaArgAr 81
   :::::|||||:::
123 .....CTGATGAA.....A 132
81 GGIYPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPR 98
133 GTCAAGAGAGCGGATGTC...CGTCAAAAAGGCGCAATGCTGTTGAAGA 179
   :::::|||||:::
98 LaArgGlyHisArgAlaProArgArgArgArgGlyAlaAlaAlaArg 114
   :::::|||||:::
180 .....CAAAAGCAATCGGCGCT..... 197
115 AlaAlaArgArgAlaProAlaAlaArgArgGlyHisGlyArgProGlyLe 131
   :::::|||||:::
198 .....GGTGTCTACGCGCGCGCTGAGCAAAATGCGCGCGATTT 237
   :::::|||||:::
131 uArgSerLeuHisValArgValHisGlyAlaAlaArgGlnGlyArgSerAsp 148
   :::::|||||:::
238 CACCGTGGGGA..... 251
148 IGRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPR 164
   :::::|||||:::
252 .....GCGGTACTTCAGTCAGT..... 269
165 ArgArgProGlyGlyProGlyAlaAlaLeuValAlaGlnLeuArgPr 181
   :::::|||||:::
270 .....CGTGAATGCGCGTTCGAGCGCAACGACGCAATTC 300
   :::::|||||:::

```

```

181 oValGlyLeuArgArgValArgAspAlaArgArgArgArgArgArg 198
301 GAGTTTGAAAGCTAGCAACCTGAGCGGCAATTCAGCGGCGAAGA 350
198 sPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPR 214
351 AGTGGCGCGCAACCTGATTCGAGTTT.....GTGACCTGGCG 391
   :::::|||||:::
215 GlyAlaArgGlyAlaGlyAspArgValGlnLeuGlyProGlyAlaAspGlyAs 231
   :::::|||||:::
392 TGGCGCGCGCGCTGTCGAGCAAAATTCGCGCGTATGCGCGCGCTT 440
   :::::|||||:::
231 rAlaHisGlyArgValProArgArgGly.....ArgProGlyAlaVal 246
   :::::|||||:::
441 .....CGCAT...CTTCGTCAATGCGATGACACCAATCGCT 476
   :::::|||||:::
246 AlaAlaProProArgHisAspGlyArgGlyLeuAspProAlaGlnPro 262
   :::::|||||:::
477 GCGTGGCGACCTTACGCTATTCAGAAAGCGCGCGAGATTCAAC 526
   :::::|||||:::
263 ArgSerHisProArgGlyLeuProArgAlaProArgArgGlyProArgAr 279
   :::::|||||:::
527 GCGCGCGCTTGGATTCAGCGCTTTCAGCGCAACGCAAAATTCATGTTGT 576
   :::::|||||:::
279 gArgAspArgGlyValAspLeuValAspArgProProAsp..... 292
   :::::|||||:::
577 AAGGCACTGGCGCAGACGTCGCGTCTGAAATGCTGCCAATCATGAAC 626
   :::::|||||:::
293 .....ArgGlyArgArgProGlyVal.....AlaGlnHisPro... 303
   :::::|||||:::
627 ACATGAATTCGCGCGCGCGCATTCGCGTTGAGTGGCAGCGCATTC 676
   :::::|||||:::
304 .....LeuArgProProAla.....ArgGluPr 311
   :::::|||||:::
677 ATTTCATGAGCGGTCGCGCGCAATTAACCGTGTGACATCAATTAT 726
   :::::|||||:::
311 oAlaHisLeuArgAlaArgArgGlyAlaGlyAlaValSerAspArgGlyP 328
   :::::|||||:::
727 CAAGATGTAATTACATTCGCGCGCTTGTTCGACAGCGCGCTGACAC 776
   :::::|||||:::
328 roArgArgArgProHis.....ArgArgAspArgArg 338
   :::::|||||:::
777 CGAGCGCGATTCGCGCTAGGTGTTCTCAGTCACAAACGCGCGCTGT 826
   :::::|||||:::
339 ArgAlaArgGlyLeuAlaArgArgGlyAlaAspProGlyAlaValProG 355
   :::::|||||:::
827 TGGCTACCGTTTGGTGGCAAGTATCGCAATTAATCGCGCGCAATTG 876
   :::::|||||:::
355 uAla..... 356
877 GTTGACACAGACAAACCGCGATTCGCGTTCGATTTGA..... 917
   :::::|||||:::
357 .....ProHisAspArgArgAlaAlaValGlnsAspArgArg 368
   :::::|||||:::
918 .....CGGCGCGATTAC.....ACAAGCGCGCGACG 943
   :::::|||||:::
369 ProGlyProLeuLeuArgArgArgArgHisArgAlaAlaArgAlaHisAr 385
   :::::|||||:::
944 ATTATTTGGAGCGCTACACCAATTCAGATTCGTTATCGAAGA...AGCG 990
   :::::|||||:::
385 gProSerGlyGlyAspProArgLeuProHisArgAlaAlaArgArgAspArg 402
   :::::|||||:::
991 CGCAGCAAAAGCTGTTGCGGTGCGTTCGCGC..... 1022
   :::::|||||:::
402 LysArgProArgAlaAlaProLeuGlyArgGlyAlaGlyArgGlyGlyG 418
   :::::|||||:::
1023 .....GCAGCGCGCAAAATTCATTCATTCACGCGCTA 1051
   :::::|||||:::
419 AspArgSerValGlyArgGlyAlaProGlyArgValAlaArgArgArgArg 435
   :::::|||||:::
1052 CAAC.....CTTCGCGCATTCCTGAAAACAAACATCTTC 1086
   :::::|||||:::
435 gArgArgGTPArgArgArgAlaProArgLeuArgProGln...GluAlaAlaG 451
   :::::|||||:::

```

```

1087 AAGTTCACACAGCCGCTCAACGCGGCGGCGCCATGTCGCGATTGG 1136
      |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
451 lYValHisProArGlyArgProGlyAla..... 465
1137 TACTACAGCGCGGTGATGCCCTTGATATCTGCCACCGCTTTTGC 1186
      |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
466 .....AlaGluArgGlnArgGlnGlyProArgArgProAlaGlySe 480
1187 GGGATTATCTCGCGCATACCGACAGCG..... 1217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 rArgArgGlyArgAlaAlaArgArgGlyArgProAlaAspArgAspA 497
1218 .....GCAGCATTTGGATTCTTGATTTGAGTGGAGCA 1247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 rGAlaAlaHisArgLeuGlyAlaArgGlyLeuAlaGlyArgGlyArg 513
1248 AGAAGA.....CTCGCTTTGTGCAGCTTGCTGCCGCGCAAT 1288
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514 ArgArgAspArgGlnProLeuArgAlaArgLeuHisLeuAlaAlaPro.. 529
1289 ACGAATACGCGCGCTGTGGCGCAAGTCT 1319
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 .ArgAlaArgAlaAlaProAlaArgArgAla 539

```

seq_name: /cgn2.6/prodata/1/aa/5B_COMB.pep:US-09-413-814-78

seq_documentation_block:

```

; Sequence 78, Application US/09413814
; Patient No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoefle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-78

```

alignment_scores:

```

Quality: 138.00 Length: 530
Ratio: 0.723 Gaps: 21
Percent Similarity: 36.038 Percent Identity: 23.019

```

alignment_block:

US-09-303-518d-125 x US-09-413-814-78 ..

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

```

9 AATCAAAAAGGTAAACCTGCGCCATCGCGGCGAGACCGAGCAACCG 58
  ::::::::::::::::::::
359 SerArgArgArgSerProAlaArgThrProGlyProArgSerProAr 375
  ::::::::::::::::::::
59 TTTCAGAGCGCGCGCATACGAAGTCCGTTGCTTG..... 98

```

```

375 gArgArgProSerProSerAlaArgSerProAlaIleTrpGluArgPro 392
99 .....CGA 101
392 roArgArgProArgAspArgArgProArgArgLeuAlaArgProAlaArg 408
102 AGAATATCGCGGTATGCGCCCTCGATGAATGCAAGAGGAGGATCGG 151
409 ArgLeuAlaArgArgAspProGlyAspValProAspProArgAlaAla 425
152 TCAAAAAGCGCAAGTCTTGATGACGCAAAAAGATCCGCGCGTG 201
425 gArgArgAlaProAlaAlaGlyAlaLeuProGlyAlaAspGlyArgGly 442
202 TTACTGCGCGCGCTTCAGCAAAATCGCGCGATTCACCGTGGCAAAA 251
442 lArgGlyProGly.....ArgArgArgAlaProArgArgGly 454
252 GCGCGTACTTCAGTCACTGCTGATTCGCGTGAAGCAACGACGAATCG 301
455 AlaAlaAlaArgValProProAspArgAlaAspProProArgArgProAl 471
302 AGTTGAAGCTACGCGACCTGAAGCGCT..... 329
471 aProAlaValAlaArgAlaAlaProAlaValAlaArgArgSerAlaArg 488
330 .....GCGAAA 335
488 lArgGlnProGlyValGlnHisProValArgGlyAlaArgAspArgPro 504
336 CTTAACGCGGAGAGAGTGGCGCGCAACTGATCCATCGCTTTGGA 385
505 AlaArgArgArgArgAlaProProGlnProValArgAspArgAlaAlaPr 521
386 CTGCGCTGCGCACCGCTCGCTTCAGCAAAATCTCGCTCATGCGCGAG 435
521 cArgGlyAlaAlaAspAspValGlnArgGlnGly..... 532
436 CGCTTGCGCATCTTCCTCAATGGATGACACCAATCGCTGCG..... 479
533 .....ArgArgAlaValProGlyArgArgAlaArgGlyProGly 545
480 .....TGCGACCGTACGGTCAATTATCAAGAGC..... CG 511
546 AlaValProAspValArgProArgAlaProArgGlyArgGlyProArgArg 562
512 CCGAGGATTTCAACCGCGCTGTGTGT.....ATTGAGCGGT 549
562 gGlyGlyLeuGlyAlaArgProArgGlyLupProArgAlaValArgProG 579
550 TTGACCGAAGCGCAAAATCCATGTTGTGAAGCACTGCGCGCAGACGTGC 599
579 lProArgAlaAlaProPro..... 585
600 GTCTGAATAATGCTGCCACATCGAAACATGATTCGCGCGCGCATC 649
585 ..... 585
650 CTGCGGTTTGAAGTGGCAGCACATTCATTTCAGAGCGGTCGCGCG 699
586 .....ArgAlaGlyHisProG 591
700 AATTAACCGTGTGACCATCAATTATATCAAGATGTAATTACATTGGCGG 749
591 l..... 591
750 TTGTTCACAGAGCGCTGTGAACACGAGCGGCTGATGCCCTAGTGTG 799
592 .....AlaProArgArgAlaArgHisArgAlaArgAlaArgAlaProArg 605
800 GTTTCAGTCAACAAACCGCGCTTCGTTACCGTTTGGGTGCGAAA 849

```

```

606 .....A 606
850 GNATGCAATTAATCTGGGGGAATGGTTGACAGCAACCGCGTAT 899
    : : : : :
606 rGValaTgArGLeuValIglYArG..... 613
900 TTCCGGTTGCGATTGTAACGGCGGATTACAAAGCGCGCGCATTTAT 949
    : : : : :
614 ..ArGLeuArGArGArGAlaArGArGAlaLeuArGArGLeuArGAlaGl 629
950 TGGACGCTACCAATCAGATTC.....CGTTATCGAA 984
    : : : : :
629 yProAlaPheProAlaAlaIglYAlaProGlYAlaValArGArGLeuArG 646
985 GAAGCGCCAGCAAGACGCTGTGG..... 1010
646 rGSeTProAlaIglYAlaValaArGArGArGArGArGArGArGAlaPro 662
1011 .....CTGGGTTGCGCGCGACGCGCGCAAAATATCTCA 1042
663 ValLeuAsPArGAlaLeuIglYArGAlaAlaProGlYPro..... 676
1043 TCACGCGTACAACTCGCGCATTTCTGAATAACAACTCTTCA...G 1089
    : : : : :
677 .ArGAlaIglYArGProArGAlaIglYAlaLysAsPValaAlaArGAla 693
1090 TTCACACAGCGCTCAAGCGCGCA..... 1115
693 IaAlaIglProArGArGArGArGAlaAlaLysProArGAsPAlaIglYAl 709
1116 .....CCGCGCCATGTCGCGGATTCGTTACTTACG 1144
710 LeuArGAlaIglYAlaSPLeuProLeuAlaLysIglYProAlaArGArGAla 726
1145 AGCGGCTGATCCCTTGATATCCCTGCGCCACCTGCTTTCG.....C 1188
    : : : : :
726 naIaArGProAla.....ProAlaAlaArGAlaArGArGLeuProArG 740
1189 GATTAAATGTCGCGCA.....TACCGACAGCGCGCGAGGCAATT 1226
    : : : : :
740 rGAlaLysArGArGArGAlaProGlYProArGArGAsPArGAlaAsPAr 756
1227 GGGTTCCTTGAAATGACAGAGAGACCTGCTTGTGACGCTTCGCT 1276
    : : : : :
757 ArGLeu.....ArGAlaPro.....AlaArGAl 765
1277 GCCCGGCAATACGAATACGCGCGCTGTGCGCAAGT 1316
    : : : : :
765 aProAsPArGLeuArGAlaArGProAsPValaArGArGAla 778

seq_name: /cgn2_6/ptodata/1/aa/6B_COMB.pep:US-09-105-537-2
seq_documentation_block:
: Sequence 2, Application US/09105537A
: Patent No. 6265202
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and plikromycin
: FILE REFERENCE: 600,438051
: CURRENT APPLICATION NUMBER: US/09/105,537A
: CURRENT FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 5215
: TYPE: PRT
: ORGANISM: Streptomyces venezuelae
US-09-105-537-2

```

```

alignment_scores:
  Quality: 105.50      Length: 601
  Ratio: 0.425        Gaps: 31
  Percent Similarity: 41.265      Percent Identity: 21.298

```

```

alignment_block:
US-09-303-518D-125 x US-09-105-537-2

```

```

Align seg 1/1 to: US-09-105-537-2 from: 1 to: 5215

```

```

31 CCCATCGCGGCGACGCGGACGCGCGCTTAC..... 63
    : : : : :
1602 ProValPheAlaAlaLeuAsPThrValTYrAlaAlaLeuAsPArgAl 1618
64 ...GACGCGCGCGCATTCAGTAAGTCCGCTTGGCGAAGATAT 108
    : : : : :
1618 sLeuAsPArGPro...LeuArGAlaLeuAlaAlaIglYAlaLysIglYAla 1634
109 .....GCGGCTATGCGCGCGCTCGATGAATGCAAGAGCGCATGCGC 150
    : : : : :
1634 sPLeuThrAlaTYrThrAlaProAlaLeuPheAlaPheGlu..... 1647
151 GTCAAAAAGCCCAAGTCTGTTGAAGACAAAAGATCCGCGCTGCT 200
    : : : : :
1648 .....ValaLeuPheArGLeuLeuGluAlaLysIglYLeuVala 1660
201 G.....TTACTGCGCGCGCTTCAGCAAAATCGCGCGATTCAC 240
    : : : : :
1660 lProAsPLeuLeuThrGluLysIleSerValIglYAlaAlaAlaAlaLys 1677
241 .....CGT 243
1677 AlaAlaIglYAlaLeuSerLeuAsPAlaAlaAlaArGLeuValThrAlaArG 1693
244 GCGGAAAAGCGGCTACTCAGTCAGTCGATTCGCGTGAAGCGCAACA 293
    : : : : :
1694 GlY.....ArGLeuMetGlnSer.....AlaArGAlaIglYAla 1705
294 CGAAATCGAATTTGAACGCTACGACGCTGAAGCGCTGCGCAACTTAAGC 343
    : : : : :
1705 aMetIleAlaValAlaGlnAlaIglYAlaIglYAlaValaLysIleSerLeu 1722
344 GCGAAGAAGTGC.....CGCAACTG 366
    : : : : :
1722 lTYrGluIglYArGValaAlaValaAlaAlaLysIglYProThrAlaVal 1738
367 ATCCATTCGCT.....TTGGGAC 386
    : : : : :
1739 ValValaSerGluAsPAlaAsPAlaAlaIglYAlaValaValaVala 1755
387 T.....GCGCTGCGGACCGCTGCTTC..... 408
    : : : : :
1755 aGluArGAlaIglYArGArGThrArGArGLeuArGAlaSerAlaPheLys 1772
408 ..... 408
1772 ePronHisMetAsPAsPValaLeuAsPAlaPheLeuArGValaAlaIglY 1788
409 .....AGCAAAATTCCTCGCGTCGATGCGGACGCGCTTCG 443
    : : : : :
1789 LeuThrPheGluIglYProArGAlaIleProValaValaSerThrValThr 1805
444 CATCTTCGTCATGCGATGAGACCAATCG..... 474
    : : : : :
1805 aLeuValThrSerGluIglYLeuThrSerProAlaTYrThrValaSPAla 1822
475 .....CTGGTGGCGAC 486
    : : : : :
1822 lLeArGArGProValaArGpHeLeuAsPAlaValaArGThrLeuAlaAlaGln 1838
487 CCTACGCTATTAACAAGAGCGCGGAGGATTTCAACGCGCGCTGCT 536
    : : : : :
1839 AsPAlaThrValaLeuValaGluIleGluProAsP.....Al 1850

```

```

1245 CGA...AGAGACCTGCGTTTGAGCAGCTGTGCCCC..... 1280
||||| .....||| |||
2127 ArgAspLysAlaAspThrLeuAlaGluIleArgLeuProAlaAlaGlnH1 2143
1281 .....GGGCAATATGGAATAGCGCCGCTTGCAGGAAGTGCT 1319
||||| :::::||| |:::::||||
2143 sglUserAlaGlyLeuPheGlyValHisProAlaLeuLeuAspAlaLa 2159

seq_name: /cgn2.6/prodate/1/liae/6B-COMB pep:US-09-156-836B-2

seq_documentation_block:
? sequence 2, Application US/09156836B
? Patent No. 6242585
? GENERAL INFORMATION:
? APPLICANT: Srivastava, Ranjana
? APPLICANT: Kumar, Deepak
? APPLICANT: Srivastava, Brahm Shanker
? TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
? FILE REFERENCE: U 011876-4
? CURRENT APPLICATION NUMBER: US/09/156,836B
? PRIOR FILING DATE: 1998-09-18
? PRIOR APPLICATION NUMBER: 08/997,897
? NUMBER OF SEQ ID NOS: 7
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 2
? LENGTH: 430
? TYPE: PRT
? ORGANISM: Mycobacterium tuberculosis
? FEATURE:
? NAME/KEY: UNSURE
? LOCATION: (4)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (6)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (20)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (29)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (54)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (64)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (69)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (89)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (99)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (114)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (119)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (129)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (159)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE
? LOCATION: (169)
? OTHER INFORMATION: amino acid has not been identified
? NAME/KEY: UNSURE

```

```

LOCATION: (182)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (185)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (219)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (259)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (291)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (323)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (339)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (349)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (356)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (366)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (410)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (421)
OTHER INFORMATION: amino acid has not been identified
US-09-156-836B-2

```

```

alignment_scores:
  Quality: 104.50      Length: 405
  Ratio: 0.741        Gaps: 22
  Percent Similarity: 34.815  Percent Identity: 23.457

```

```

alignment_block:
US-09-303-518D-125 x US-09-156-836B-2 ..

```

```

Align seg 1/1 to: US-09-156-836B-2 from: 1 to: 430

```

```

102 AGAATATCGCGTATCGG.....CCCCGATGAAGCAAG 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 ArgTyrCysArgTyrProGlySerArgTyrProSerSerArgCysProSe 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 AAGCGATCGCGTCAAAAAAGGCCAAGTCTGTTGAAGCAAAAA... 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 rleuhArgCysArgArgPheArgCysProArg**ArgCysGlnArgTyrT 165
186 .....GAATCCGGCGCTGCTTTACTGCGCGCGCTTCAGGGA 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 rPcysProasn**ThreGlyArgCysCysArgCysProSerSerSerArg 181
225 AATCGCGCGATTCACCGCGCGGCAAAAGCGCTACTTCAGTCAGTCGTA 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 ***GlnTyr***SerProAlaGlyCysArgArgThrAlaArgCysArgCy 198
275 TTGCGCTTGAAGCAACGCAAAATCGAGTTTGAACGCTACGACCTGAA 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 scCysCysArgCysTrpArgThrArg..... 207
325 GCGCTGGCAACTTAAGCGCGCAAGAGTGGCGCCGCAACCTGATCCATC 374

```

```

207 ..... 207
375 CGGTTGTGACTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 .....CysCys.....CysCy 211
425 TCGATGCCGAGCGGTTGCCATCTTCGTAATGGATGACACCAATCCG 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
211 sArgCys.....TrpGlnSerL 217
475 CTGCG...TGCGGACCTAGCGTCATTATCAAGA...AGCGCGGAGGA 518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 euGly***SerArgProArgSerArgSerArgCysSerArgArgArg 233
519 TTTCAACGCGCGCGCTGTGTGATTGAGCGGTTTGACCAACGCAAAATCC 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 PheGlnAsnArg..... 237
569 ATGTTGTAGGACAGCTGCGGCGGAGAGTGGCGTGAATAATGTCGCAAC 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 .....CysCysArgS 241
619 ATCGAABACATGATGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 erArgGlyPheArgIleArg.....CysCysSerPhe..... 251
669 GCACATTCATTCATCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 .....ProGlyPheArg.....AsnArg.....Hi 258
719 TCATATTATCAGATGATTAATACCATGGCGGCTTTGTTGCAACAGGCGGT 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 s***IleLeuArgCys.....PheHisCysArgTyr***S 270
769 CTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 erCysArgArgCysArg...CysProArgCysPheGlyCysArgGlyCys 285
813 .....CAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 ArgCysGlnGlyCys***SerHisArgArgPheArgCysArgGlyCysCy 302
857 AAT...TACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 saenCysArgCysTrpArgCys..... 309
904 GGTTCGATATTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
310 .....ArgGlyCysSerArgArgProGlyLeuProGly 320
954 .....ACGTAACCAATCAGATTTCG 976
321 ArgAsp***ArgProValGlyHisArgGlyIleProThrCysCysPheAr 337
977 TTTTCGAAGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 gCys***ArgSerProArgSerArgProAlaLeu***TyrProProGly 354
1018 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1067
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 erCys***ThrAsnProIleArgCysCysProSer***SerArgProIle 370
1068 CTGGAANAACAACTCTTCAGTTCAACACGCGGCGGCGGCGGCGGCGG 1117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 ProAlaArgProArgLeuPro.....GlyArgSerTyrArgTrpPr 384
1118 G.....CGCATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 oProThrLysSerGlyArgSerGlnAsnCysTrpPheHisArgSerSerGly 400
1153 ATGCCCTTGATATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 .....SerArgThr 403

```

1203 CGATACCGACAGCGC 1217
|||||
404 ArgTyraIghIsarg 408

seq_name: /cgn2_6/ptodata/1/laa/6B.COMB.pep:US-09-036-987A-6

seq_documentation_block:

Sequence 6, Application US/09036987A

Patent No. 6143526

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.

APPLICANT: Broughton, Mary C.

APPLICANT: Crawford, Kathryn P.

APPLICANT: Madduri, Krishnamurthy

APPLICANT: Merlo, Donald J.

APPLICANT: Treadway, Patli J.

APPLICANT: Turner, Jan R.

APPLICANT: Waldron, Clive

TITLE OF INVENTION: Biosynthetic Genes for Spinosyn Insecticide

TITLE OF INVENTION: Production

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dow Agrosciences LLC Patent Department

STREET: 9330 Zionsville Road

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/036,987A

FILING DATE: 09-MAR-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stuart, Donald R.

REGISTRATION NUMBER: 28,479

REFERENCE/DOCKET NUMBER: 50,608

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317)337-4816

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 5588 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-036-987A-6

Alignment_scores:

Quality: 104.50

Ratio: 0.469

Percent Similarity: 41.220

Percent Identity: 20.518

Length: 541

Gaps: 26

Alignment_block:

US-09-303-518D-125 x US-09-036-987A-6

Align seg 1/1 to: US-09-036-987A-6 from: 1 to: 5588

22 CTAAACCTGCC.....ATCGCGGACAGCCGAG..... 51

|||||
2424 LeuSerLeuProAspAlaAlaArgMetValAlaGlyArgAlaArgLeu 2440

52 .CAAGCGGTTTACAGCGCGCCGCCATACCGAAGTCGGCTTGGCG 100

|||||
2440 tGlnAlaLeuProSerIyGlyAlaMetLeuAlaValAla...AlaGlyG 2456

101 AAGAAATGCGCGTATGCGCCCTCGATGAAAGTCAAGAGGCGATGCC 150

|||||
2456 IuGluGln.....LeuArgProLeu.....AlaAspArg 2466

151 GTCAAAAAGGCCAAGTGGCTGTTGAGACACAAAAAGAAATCCGGCGTGT 200

|||||
2467 ValAspGlyAlaGlyLe.....AlaAl 2474

201 GTTACTGCGCGCCGCTTACGCCAAATCGCCGATTCACCGTGGCGAA 250

|||||
2474 aValAsnAlaProGluSer.....ValValLeuSerGlyAspA 2487

251 AGCGCTACTTACATGATCGTGTATCCGCTTGAAGGACACAC..... 294

|||||
2487 rGluValLeuAspAspIleAlaGlyAlaLeuAspGlyGlnGlyLeArg 2503

294

2504 TrpArgArgLeuArgValSerHisAlaPheHisSerTyrArgMetAspPr 2520

295

2520 oMetLeuGlnIuPheAlaGlyIleAlaArgSerValAspTyrArgArg 2537

314 ACGACCTGAGCGCGTGGCAACTTAAGCGCGAA..... 348

2537 LysAspLeuProValValSerThrLeuThrGlyGlyLeuAspThrAlaGly 2553

349

2554 ValMetAlaThrProGlyTyrTrpValArgIleValArgIuProValArg 2570

358

2570 gPheAlaAspGlyAlaArgValAlaLeuAlaGlnGlnGlyValAlaThrIleP 2587

392 TCGCACCCGTCGCTTACGAAATTCCTGCGCGATTCGCGCGCGCTTC 441

|||||
2587 heGluLeuGlyProAspAlaThrLeuSerAlaLeuIleProAspCysHis 2603

442 GCCATCTTGCATATGCGATGAGACACCAATCCGCTGCTGCGACCTAC 491

|||||
2604 SerTrpAlaAspGlnAlaMetProIleProMetLeuArgLysAspArgThr 2620

492 G.....GTCAATTATCAAGAAGCCGCCGAGATTTAAAGCGGCC 532

|||||
2620 rGluThrGluThrValValAlaAlaValAlaArgAlaHisThrArgGly 2636

533 TGTGGTATGAGCCGTTTACCGAAGCAAAATTCATGTTGTAAAGCA 582

2636

583 GCTGGCGAGACGTCGCGCTGAAATGCTGCCAATCGAATCGAATGACATGA 632

|||||
2637

633 ATTCGCGCGCCGCGATCGCGGTTGAGTGGACAGACCAATTCATTCTTA 682

|||||
2644

683 TCGAGCGGTCGCGCGCAATAAACCGTGTG.....ACCATCAATTAT 726

|||||
2656 rGlnTrpAlaPheGlnArgGlnArgTyrTrpLeuGluThrSerAspTyr 2672

727 CAAGATGATTAATACATGCGCGCTTGTGTTGCAACAGCGCGTCTGAACAC 776

|||||
2673 GlyAspValThrGlyIleGlyLeuAlaAlaAlaGly..... 2684

777 CGAAGCGGTATGCTAGTGTGTTCTCAAGTCAACAAACGCGCGCTCT 826

2684

827 TCGGTACCGTTTGGTGGAAAGTATCGAAATTAATCGCGCGCAATTC 876

|||||
2684

[illegible]

```

seq_name: /cgn2_6/prodata/1/iaa/6A_COMB pep:US-09-335-409-5
seq_documentation_block:
? Sequence 5, Application US/09335409
? Patent No. 6121029
? GENERAL INFORMATION:
? APPLICANT: Schnupp, Thomas
? APPLICANT: Ligon, James
? APPLICANT: Molnar, Istvan
? APPLICANT: Zikler, Ross
? APPLICANT: Cyr, Devon
? APPLICANT: Goerlach, Joern
? TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
? FILE REFERENCE: 4-30582A
? CURRENT APPLICATION NUMBER: US/09/335,409
? CURRENT FILING DATE: 1999-06-17
? NUMBER OF SEQ ID NOS: 30
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 5
? LENGTH: 7257
? TYPE: prt
? ORGANISM: Sorangium cellulosum
US-09-335-409-5

```

alignment_scores:	
Quality:	103.50
Ratio:	0.452
Percent Similarity:	42.329
alignment_block:	
	Length: 541
	Gaps: 27
	Percent Identity: 21.442

US-09-303-518D-125 x US-09-335-409-5 . .

Align seg 1/1 to: US-09-335-409-5 from: 1 to: 7257

```

55 GCGGTTACACGCGCCGCCATTACCAGGATGCGCTTGCGGCAA... 102
6354 AlaserhearsprijuglnalaprrothrservValahnlseuSierle 6370
103 .GAATATGCCGGATATGCGCCGCCCTTCGATGAATCAAGAAAGCGATGCCG 151
6370 uasrglUatrglUatVleuAspAlaAspAlaIarProthrsAlaAspAlal 6387
152 TCAAAAAAGGC.....CAATGCGTGTTCATAACACAA 183
6387 euglUuglUserleValaIargglUcysAspSerValleuIrrhValgln 6403
184 AACAAATCCGGCGGTGGTGTTCCTACTGCGCCG.....GCTTC 218
6404 AlalValalagllUatIarPhreatgAspRproIarProIarUleuTrpleuValTh 6420
219 AGGCAAAATCGCGCGCATTCACCGGCGGCAAAAGCGGTACTGATCGTAG 268
6420 rarglUatIarglnalalIeeglUalagllUAspAlaSerValalaglnAlar 6437
269 TC.....GTGATTCGCGTTGAA..... 285
6437 rUleuUeuglUleuUglUatrgUalIlealaleuUglnlSialglUleuIarg 6453
286 .....GGCAAGCGCAAAATCGAGTTTGAAAGC 311
6454 CysAlaIatrglIleAspRleuAspProIalargIargAspIuglUatVleuAspI 6470
312 CTACGCCACTGAAAGCGCTGCGCAAACTTAAGCGGCGGCAAAAGAAATCGCGCGCA 361
6470 UleuUleuUaIeUleuUeulAlaAspAlaIagllUglUatVAlaIaIarPheA 6487
362 ACCTGATCCAAATCGCGTGTTCGAGTCCGCGGCGCGCACCGCGCGTTGACG 411
6487 rgglUglUglUatrgUatVAlaIarIargUleuValaIargUleuProglUthr 6503
412 .....AAATTCCTCGCGCGTGGATGCCAGCGCGTTGCGCATTT 449

```

```

6504 AspCysArgGluTyrIleGluProAlaGluGluTyrProPheArgLeuGlu 6520
450 CGTCAAT.....GCGATGGACA 466
6520 uileAspGlySerGlyValLeuAspPheValLeuArgAlaThrGlu 6537
467 CCATCCGGCTGGCCGACCCCTACGCTATTATCAAGAGCCGCCGAG 516
6537 rGArgProProGlyProGlyGluValGluIleAlaValAlaAla... 6552
517 GATTTCAAACGGCGCTGTGGTATGACCCGTTGACCGAAGCAAAAT 566
6553 .....GlyLeuAsnPheLeuAspValMetArgAlaMetGlyI 6565
567 CCAT.....GTTGTAG. 579
6565 eTyrProGlyProGlyAspGlyProValAlaLeuGlyAlaGluCysSerG 6582
580 .....GCACTGGCGGACGCTGCTCT.....GAA 606
6582 lYArgIleValAlaMetGlyGluGlyValGluSerLeuArgIleGlyGln 6598
607 AATGCGCCCAACATCGAACAACATGATCGGC..... 639
6599 AspValValAlaValAlaProPheSerPheGlyThrHisValThrIleAs 6615
640 .....GCGCCGATCCCTCCGTTGAGTGGCAGCCGACATTC 676
6615 pAlaArgMetLeuAlaProArgProAlaIleuThrAlaAlaGlnAla 6632
677 ATTTCATGACCGCGTGGCGCATTAACCGTGTGACATCAATTAT 726
6632 laAlaLeu...ProVal...AlaPheMetThrAlaTrp..... 6642
727 CAAAGTAATTAATCAATGGCCGTTGTTGCAACAGCCGCTGTAACAC 776
6643 TyrGlyLeuValHisLeuGlyArgLeuArgAla.....G 6654
777 CGACCGCGTAT.....GCCCTAGGTGTCTCAAGTCACAAAC 817
6654 yGluArgValLeuIleHisSerAlaThrGlyGlyThrGlyLeuAlaVal 6671
818 CGCGCTCTTCCGTCACGCTTTGGTGGCAAGTA..... 852
6671 alGlnIleAlaArgHis...LeuGlyAlaGluIlePheAlaThrAlaGly 6686
852 ..... 852
6687 ThrProGluLysArgAlaTrpLeuArgGluGlnGlyIleAlaHisValMe 6703
853 .....TCGCAAT.....ACTGCG 868
6703 tAspSerArgSerLeuAspPheAlaGluGlnValLeuAlaIleThrLys 6720
869 GCGAATGGTGGACACAGACAAACCGCGTATTCGCGTGGATTGAAAC 918
6720 yGluGlyValAlaAspValValLeuAsnSerLeuSerGlyAlaAlaIleasp 6736
919 GCGCGGATTACAAAGCGCGCAGATTTTGGAGCGTACACAAATCA 968
6737 AlaSerLeuSer..... 6740
969 GATTTCGTTATCGAAGAGCCGACGAAAGACGCTTGGCGTGGGTTG 1018
6741 .....ThrLeuValProAspGlyArgPheIleGluLeu..... 6751
1019 CGCGCCAGCCGACAAATACCTACGACGCTACAAACCTCGCGCATTC 1068
6752 ...GlyLysThrAspIleTyrAlaAspArgSerLeuGlyLeuAlaHisPhe 6767
1069 CTGAAACAAACCTCTCAAGTTCACACAGCCGTCACAGCGCGCGACGC 1118

```

```

6768 ArgLysSerLeuSerTyr.....SerAlaVal.....AspLe 6778
1119 CGCCATGTCGCCGATGTACTTACGAGCGCGT..... 1152
6778 uAlaGlyLeuAlaValAlaArgArgProGluArgValAlaIleAlaLeuAla 6795
1153 .....ATGCCCTTGAT 1164
6795 lValValAlaAspLeuLeuAlaArgGlyAlaLeuGlnProLeuProValGlu 6811
1165 ATCTGCGCCACCCCTGCTTTCGCGATTTATCGTCGGCGATACCGACAG 1214
6812 lIlePheProLeuSerArgAlaAlaAspAlaPhe...ArgLysMetAlaGln 6827
1215 CGCGCAGCATGTGGTGTCTTGAAATTGACAGAAAGAACCTCGCTTGT 1264
6828 AlAGlnHisLeuGlyLysLeuValLeuAlaLeuGlnLysProAsp...Va 6843
1265 GCAGCTTCGCTGCGCGGCA 1286
6843 lArgIleArgValProGlyGln 6850

```

```

seq_name: /cgn2_6/ptodata/1/laa/5b_COMB.pep:US-09-568-102-5
seq_documentation_block:
; Sequence 5, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: prt
; ORGANISM: Sorangium cellulosum
; US-09-568-102-5

```

```

alignment_scores:
  Quality: 103.50      Length: 541
  Ratio: 0.452        Gaps: 27
  Percent Similarity: 42.329  Percent Identity: 21.442

```

```

alignment_block:
US-09-303-518D-125 x US-09-568-102-5 ..
Align seg 1/1 to: US-09-568-102-5 from: 1 to: 7257

```

```

55 GCCGTTACGACGCGCCGCTTACGAGTGCCTTGCGCGAA... 102
   ||| ..... ||| ..... ||| ..... ||| .....
6354 AlaserPheAspGlyGlnAlaProThrSerValValHisLeuIleYserLe 6370
103 GAATATCGCGTATGCGCCCTCGATGAAGTCAAGAGCGGATCGCG 151
   ||| ..... ||| ..... ||| ..... ||| .....
6370 uAspGluArgGlyValLeuAspAlaAspAlaProPheAspAlaAla 6387
152 TCAAAAAAGC.....CAAGTCTGTTGAAGACAA 183
   ..... ||| ..... ||| ..... ||| .....
6387 euGluGluSerLeuValArgGlyCysAspSerValLeuThrPrrValGln 6403
184 AAGAATCGCGCGTGTCTTACTGCGCG.....GCTTC 218
   ||| ..... ||| ..... ||| .....

```

```

6404 AlaValAlaGlyAlaGlyPheArgAspProProArgLeuThrIleValTh 6420
219 AGGCAAAATCGCGGATTCACCTGGCGAAACCGGCTACTTCAGTACG 268
6420 rArgGlyAlaGlnAlaIleGlyAlaGlyAspValSerValAlaGlnAla 6437
269 TC.....GTGATTCGGTTGAA..... 285
6437 rOLEuLeuGlyLeuGlyArgValIleAlaLeuGlnHisAlaGluLeuArg 6453
286 .....GGCAACGACGAAATCGAGTTTGACG 311
6454 CysAlaArgIleAspLeuAspProAlaArgArgAspGlyGluValAspG 6470
312 CTACGCACTGAACGCTGGCAAACTTAACGGCGAAGAGTCCGCCGCA 361
6470 uLeuLeuAlaGluLeuLeuAlaAspAspAlaGluGluGluValAlaPhe 6487
362 ACCTGATCCAAATCCGTTTGTGACGCTGGCGACCCGCTCCGTTACG 411
6487 rGgGlyGlyArgArgValAlaArgLeuValArgArgLeuProGluThr 6503
412 .....AAATCTCGCCGTCGATGGCGACCGCTTCCCATCTT 449
6504 AspCysArgGlyLysIleGluProAlaGluGlyArgProPheArgLeu 6520
450 CGTCGAT.....GGATGGACA.466
6520 uIleAspGlySerGlyValLeuAspLeuValLeuAlaArgAlaThrGln 6537
467 CCATCGCGCTGGCTGGCGACCTACGCTATTAACAGACCGCCGAG 516
6537 rGArgProProGlyProGlyGluValGluIleAlaValGluValAla... 6552
517 GATTTCAACGCGCCGTTGTGATGAGCGTTTGACCGAAGCAAAAT 566
6553 .....GlyLeuAsnPheLeuAspValMetArgAlaMetGlyI 6565
567 CCAT.....GTTTGTAG. 579
6565 eIyTrProGlyProGlyAspGlyProValAlaLeuGlyAlaGluCysSer 6582
580 .....GCAGCTGGCGACAGCTGGCGCT.....GAA 606
6582 yArgGlyLeuAlaMetGlyGluGlyValGluSerLeuArgIleGlyGln 6598
607 AATGCTGCCAATCGAAACACATGAATTCGCG..... 639
6599 AspValAlaValAlaProPheSerPheGlyThrHisValThrIleAs 6615
640 .....GGCCGCACTCGCGGTTTGAGTGGCGACGACATTC 676
6615 pAlaArgMetLeuAlaProArgProAlaIleLeuThrAlaGlnAla 6632
677 ATTTCAATCGACCGGCTGGCGGATTAACCGTGGACCATCAATTA 726
6632 lAlaLeu...ProVal...AlaPheMetThrAlaTrp..... 6642
727 CAAGATGAATTAACATGGCGCTTTGTTCAACAGCCGCTGAAAC 776
6643 TyrGlyLeuValHisLeuGlyArgLeuArgAla.....GI 6654
777 CGAGCGCTGAT.....GCCCTAGGTGTTCTCAAGTCAACAAC 817
6654 yGluArgValLeuIleHisSerAlaThrGlyGlyThrGlyLeuAlaVal 6671
818 CGCGCGCTCTGCTACCGCTTTGGGTGCGAAGTA..... 852
6671 aGlnIleAlaArgHis...LeuGlyAlaGluIlePheAlaThrAlaGly 6686
852 ..... 852
6687 ThrProGlyLysArgAlaTrpLeuArgGluGlnGlyIleAlaHisValMe 6703

```

```

853 .....TCGCAAAAT.....ACTGCG 868
6703 tAspSerArgSerLeuAspPheAlaGluGlnValLeuAlaAlaThrIlysg 6720
869 GCGAATTTGTTGACACAGACACCGCGTGAATTCGCGTTCCGATTTGAC 918
6720 yGluGlyValAspValValLeuAsnSerLeuSerLeuGlyAlaAlaIleAs 6736
919 GCGCGATTCACACAGCGCGCACGATTTTGGACGCTTACCACATCA 968
6737 AlasLeuSer..... 6740
969 GATTTCCGTTATCGAAGAGCCGACGAAAGAGCTGTCGCGTGGGTTG 1018
6741 ....ThrLeuValProAspGlyArgPheIleGluLeu..... 6751
1019 GCGCGACGCGCGACAAATTAATCTCATCAGCGGTACAACTCGCCATTTG 1068
6752 ..GlyLysThrAspIleTyrAlaAspArgSerLeuGlyLeuAlaHisPhe 6767
1069 CTGAACAACTCTTCAAGTTCAACACAGCGGTACACGCGCGACCG 1118
6768 ArgLysSerLeuSerTyr.....SerAlaVal.....AspLe 6778
1119 CGCCATGTGTCGATTTGATTACAGCGCGCTG..... 1152
6778 uAlaGlyLeuAlaValAlaArgProGluArgValAlaAlaLeuLeuAlaG 6795
1153 .....ATGCCCTTGAT 1164
6795 lValValAspLeuLeuAlaArgIlyAlaLeuGlnProLeuProValGlu 6811
1165 ATCTTCCGACCCCTGCTTTGGCGATTTAATCGCGGATTAACGACAG 1214
6812 lIlePheProLeuSerArgAlaAlaAspAlaPhe..ArgLysMetAlaGln 6827
1215 CGCGACGAGGATTTGCTTGGATTTGACAGACAGACCTCGCTTGT 1264
6828 AlAGlnHisLeuGlyLysLeuValAlaLeuAlaLeuGlnAspProAs...Va 6843
1265 GCAGCTTCGCTGCGCGGCA 1286
6843 lArgIleArgValProGlyGlu 6850

```

seq_name: /cgn2_6/ptodata/1/aa/6b_COMB.pep:US-09-567-969-5

seq_documentation_block:

Sequence 5, Application US/09567969

Patent No. 6355457

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/567, 969

CURRENT FILING DATE: 2000-05-10

PRIOR APPLICATION NUMBER: 09/335,409

PRIOR FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 7257

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-567-969-5

alignment_scores:

Quality: 103.50 Length: 541
Ratio: 0.452 Gaps: 27
Percent Similarity: 42.329 Percent Identity: 21.442

alignment_block:

US-09-303-518D-125 x US-09-567-969-5

Align seg 1/1 to: US-09-567-969-5 from: 1 to: 7257

```

55 GCCGTTACGAGCGCCGCGCATTTACCGAAGTCGCGTTGCGCGAA..102
||| :|||:||||| ||| :|||:||||| :|||:|||||
6354 Alaserhelspdlglnalaprothrservalhisleuglyserle 6370
103 .GAATATCGCGTATGCGCCCTCGATGAAGTCAGGAAGCGCATGCC 151
||| :|||:||||| ||| :|||:||||| :|||:|||||
6370 uaspglarsgllyallleuaspalaasralaprotheaspalaal 6387
152 TCAGAAAAGGCG.....CAGTGTCTTGAAGCAAA 183
:|||||:|||||:|||||:|||||:|||||:|||||
6387 euglugluserleuvalargllycysaspservalleutprthvalgln 6403
184 AAGATCCGCGCTGTGTCTTACTGCGCG.....GCTTC 218
|||||:|||||:|||||:|||||:|||||:|||||
6404 AlavalalaglialaglypheargAspproproargleutrpheualth 6420
219 AGCGAAATCCCGCGCATTCACCGTGGCGAAGCGGCTACTTCAGTCAG 268
|||||:|||||:|||||:|||||:|||||:|||||
6420 rargglialaglnalailleuglyalaglyaspservalalaglnlap 6437
269 TC.....GTGATTGCGCTTGAA..... 285
|||||:|||||:|||||:|||||:|||||:|||||
6437 roleuleugllyleugllyargvalillealaleuglunhisalaglualeuarg 6453
286 .....GGCAGACGCAAAATCGAGTTTGACG 311
:|||||:|||||:|||||:|||||:|||||:|||||
6454 Cysalalargileaspleuaspaleuaspaleuargargllyuvalaspdl 6470
312 CTACCGACCTGACGCGCTGCGCAACTTAAGCGCGGAGAGAGTGGCGCGCA 361
|||||:|||||:|||||:|||||:|||||:|||||
6470 uleuleualaglualeuaspaleuaspaleuaspaleuaspaleuaspaleu 6487
362 ACCTGATCCAAATCCGCTGTGTGACTGCGTGGCGACCGCTCCGTTACAGC 411
||| :|||:||||| ||| :|||:||||| :|||:|||||
6487 rgglygllygluargargvalalalargleuvalalargargleuprogliuthr 6503
412 .....AAATTCCTGCGCTGATGCGCGAGCGGTCGCGCATCTT 449
|||||:|||||:|||||:|||||:|||||:|||||
6504 aspcysarggllyslleugluproalagluargleuprotheargleuagl 6520
450 CGTCAAT.....GCGATGACGA 466
:|||||:|||||:|||||:|||||:|||||:|||||
6520 uilleaspdlsergllyallleuaspaleuaspaleuaspaleuaspaleu 6537
467 CCAATTCGCGTGGCTGCGCGACCTAGCATATTCAGAAAGCGCGCGAG 516
||| :|||:||||| ||| :|||:||||| :|||:|||||
6537 rgarprothroglyprogllygluvaligulllealavalaglnalala... 6552
517 GATTTCAAACGCGCGCTGTGTGATGAGCCGTTGACCGAAGCGCAAAAT 566
|||||:|||||:|||||:|||||:|||||:|||||
6553 .....Glyleuasnpheluspsvalmetargalalamegllyl 6565
567 CCAT.....GTTTGAAG 579
:|||||:|||||:|||||:|||||:|||||:|||||
6565 eyurprogllyprogllyaspglyprovalalaleuagllyalaglycysarg 6582
580 .....GCAGCTGGCGACGAGCGTCCGCT.....GAA 606
|||||:|||||:|||||:|||||:|||||:|||||
6582 lylargllyalalamegllygluvalaglyserleuargllyleugln 6598
607 AATGCTCCACATCGAAGACATGAATTCGCG..... 639
:|||||:|||||:|||||:|||||:|||||:|||||
6599 Aspsvalvalalavalalaprotheserphaglythrhisvalthrleas 6615

```

```

640 .....GCGCGCATCCGCGCGTGTGATGAGCGACATTC 676
:|||||:|||||:|||||:|||||:|||||:|||||
6615 palargmetleualapropargproalalaleuthralalaglnalaa 6632
677 ATTTCATGAGCGCGCGCGCGCATTAACCGTGTGACATCATATTAT 726
||| :|||:||||| ||| :|||:||||| :|||:|||||
6632 laalaleu...Proval...Alarhemethrallatrp..... 6642
727 CAGATGTAAATTACCATGCGCGTGTGTGACAGCGCGCTGAAAC 776
||| :|||:||||| ||| :|||:||||| :|||:|||||
6643 tyrglyleuvalnhsleugllyargleuargla.....G 6654
777 CGAGCGCGTGAAT.....GCGCTAGGTGTCTCAATCAACAAAC 817
|||||:|||||:|||||:|||||:|||||:|||||
6654 ygluarglyalleuilehisseralathrglyglythrleuvalalav 6671
818 CGCGCTGTGCGTACCGCTTGTGGTGGCAAGTA..... 852
:|||||:|||||:|||||:|||||:|||||:|||||
6671 alglillealarghis...leuglyalagluillephealathalagly 6686
852 ..... 852
6687 thrprogllyulysargalatripleuargllyuglnlylealhisvalme 6703
853 .....TCGCAATT.....ACTGGCG 868
6703 laspsertargserleuaspaleuaspaleuaspaleuaspaleuaspaleu 6720
869 GCGAATTTGTGTGACAGACAGACCGCGTGTGCTGCTGCTGATTAAC 918
|||||:|||||:|||||:|||||:|||||:|||||
6720 lygluylvalaspvalvalleuasnsertleusergllyalalaleasp 6736
919 GCGCGCATTTACACAGCGCGACGATTTGTTGGACGCTACACAAATCA 968
:|||||:|||||:|||||:|||||:|||||:|||||
6737 Alaserleuser..... 6740
969 GATTTCCGTTATGAGAAGCGCGCGCAAGAGCGTTCGCGTGAGTTG 1018
:|||||:|||||:|||||:|||||:|||||:|||||
6741 ....ThrleuvalProaspdllyargpheilleuglu... 6751
1019 CGCGCGAGCGCGCAAAATCTCCATCAGCGGTACAAACCTCGCGCATTTG 1068
||| :|||:||||| ||| :|||:||||| :|||:|||||
6752 .glylustrlrarpllelyrhalasprargserleuagllyleuallhsphe 6767
1069 CTGAATAACAACTTTCAAGTTCAACACAGCGGTCAACGGCGGACCG 1118
||| :|||:||||| ||| :|||:||||| :|||:|||||
6768 Arglyserleuserlyr.....Serlaval...Asple 6778
1119 CGGCATGTGTCGATGCTTACGAGCGCGT..... 1152
|||||:|||||:|||||:|||||:|||||:|||||
6778 ualaglyleuvalalargargprogluargvalalalaleuvalag 6795
1153 .....ATGCCCTTGAT 1164
6795 luvalalaspaleuvalalargllyalaleuaglproleuvaligln 6811
1165 ATCTGCGCCACCGCTGTTGCGCGATTTAATGCTGCGCATCCAGACAG 1214
|||||:|||||:|||||:|||||:|||||:|||||
6812 illephroleuserargalalalaspalalphe...Arglysmetalagln 6827
1215 CGCGCAGCATTTGGGTTGCTTGAATGAGACAGAAAGACCTGCTTTGT 1264
:|||||:|||||:|||||:|||||:|||||:|||||
6828 Alaglnhisleugllylsleuvalleuallaleuaglunaspasps...Va 6843
1265 GCAGCTTCGCTGCGCGCGCA 1286
|||||:|||||:|||||:|||||:|||||:|||||
6843 larglileargvalprogllygln 6850
seq_name: /cgn2_6/prodata/1/laa/6b.COMB.pep:US-09-568-480-5
seq_documentation_block:
; Sequence 5, Application US/09568480
; Patent No. 6355458

```

```

1  GENERAL INFORMATION:
2  APPLICANT: Schupp, Thomas
3  APPLICANT: Ligon, James
4  APPLICANT: Molnar, Istvan
5  APPLICANT: Zinke, Ross
6  APPLICANT: Cyr, Devon
7  APPLICANT: Goerlach, Joern
8  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
9  FILE REFERENCE: 4-30582A
10 CURRENT APPLICATION NUMBER: US/09/568,480
11 CURRENT FILING DATE: 2000-05-10
12 PRIOR APPLICATION NUMBER: 09/335,409
13 PRIOR FILING DATE: 1999-06-17
14 NUMBER OF SEQ ID NOS: 30
15 SOFTWARE: PatentIn Ver. 2.0
16 SEQ ID NO 5
17 LENGTH: 7257
18 TYPE: PRT
19 ORGANISM: Sorangium cellulosum
20 US-09-568-480-5

```

```
alignment_scores:
  Quality: 103.50
  Ratio: 0.452
  Percent Similarity: 42.329
  Length: 541
  Gaps: 27
  Percent Identity: 21.442
```

alignment_block:

US-09-303-518D-125 x US-09-568-480-5

Align seg 1/1 to: US-09-568-480-5 from: 1 to: 7257

[illegible][illegible]

6778 ualaglyleualvalargargprogluargvalalaaleuleualag 6795
 1153 ATGCCCTTGAT 1164
 6795 luvalvalaspleuleualarglylaleuenglproleuprovalglu 6811
 1165 ATCTGCCACCTGCTTTGGCGATTATTCGTCGGGATCCGACAG 1214
 6812 illephreulersearargalalaaspalaphe..Arglysmetlaagin 6827
 1215 CGCGCAGCATTTGGTTCCTTGAATTCGACGACGAGACGTCGCTTGT 1264
 6828 AlaglnhisleuglylsleuvalleualaleuGlusprohsp...Va 6843
 1265 GCAGCTTCGTCTGCCCGCGCA 1286
 6843 larginleargvalprogllyglu 6850

seq_name: /cgn2_6/ptodata/1/laa/6B_COMB.pep:US-09-568-486-5

seq_documentation_block:

; Sequence 5, Application US/09568486
 ; Patent No. 635459
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligou, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; CURRENT APPLICATION NUMBER: US/09/568,486
 ; PRIOR FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 7257
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 ; US-09-568-486-5

alignment_scores:

Quality: 103.50 Length: 541
 Ratio: 0.452 Gaps: 27
 Percent Similarity: 42.329 Percent Identity: 21.442

alignment_block:

US-09-303-518d-125 x US-09-568-486-5 ..

Align seg 1/1 to: US-09-568-486-5 from: 1 to: 7257

55 GCCGTTTACGAGCGCGCGCATTAACGAGTCGGCTTGGCCGAA.. 102
 6354 Alaserpheaspgllyglinalaprothrservalvalhisleuglyserle 6370
 103 .GAATATGCCGTATGCCGCTCGATGAAGTCAAGGAGCGCATGCCG 151
 6370 unspluarlglyvalleuasplalaaspalaproheaspalasal 6387
 152 TCAAAAAAGC.....CAAGTGTCTTTGAAGACAAA 183
 6387 eucluelserleuvalarglycysaspservalleutripthrValgln 6403
 184 AAGATCCGCGCGCTGTTTACTCGCGC.....GCTTC 218
 6404 AlavalalaglylaleuvalarglycysaspservalleutripthrValgln 6420
 219 AGGCAAAATCGCGCGATTCACCGGCGGAAAGCGCTACTTCAGTCAG 268

6420 rangglyalaglnAlaleglylaleuvalaspservalvalalaglnlap 6437
 269 TC.....GTGATTCGCTTGAA..... 285
 6437 roleuleuengllyleuarglyvalilealaleuenglhisalagluuarg 6453
 286GGCAACGACGAATCGATTCGAACG 311
 6454 Cysalarglyleaspleuasproalargargarglyluvalaspl 6470
 312 CTACGACCTGAAGCGCTGCAACTTAAGCGCGCAAGATGCGCGCA 361
 6470 uleuleuAlagluuvalleualaspsalagluuvalalapha 6487
 362 ACCTGATCCAAATCCGCTTGTGACTCGCGACCGCGCTGCTTACG 411
 6487 rglylglyluargargvalAlaargleuvalargleuproglu 6503
 412AAATTCCTGCGCTGATGCGGACCGCTTGCATCTT 449
 6504 AspCyaharglylustrilegluproalagluargprophargleu 6520
 450 GGTCAAT.....GGGATGGACA 466
 6520 uileaspglyserglyvalleuaspspleuvalleuargalatrhlua 6537
 467 CCAATCCGCTGCTGCGCGACCTACGCTATTCAGAAAGCCGCGAG 516
 6537 rgargproproglprogllygluvalgluilevalalalua 6552
 517 GATTTCAAACGCGCTGTTGATTGACCGCTTTCAGCAACGCAAAAT 566
 6553glyleuasnphelusvalmetlarylaleu 6565
 567 CCAT.....GTTTGAAG. 579
 6565 etytrproglprogllyasplprovalaleuGllyalaglycysserg 6582
 580GCAGTGGCGCAGACGTCGCTCT.....GAA 606
 6582 lyargilevalAlametglylgluvalgluuleuarglyleuglyln 6598
 607 AATGCTGCCAATCGCAACACATGATTCGCG..... 639
 6599 AspvalvalAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 6615
 640GGCCGATCCTGCGGCTTGAAGTGGCAGCAGCATTC 676
 6615 palargmetleualaproargproalalaleuThrAlaAlaAla 6632
 677 ATTTCATCGAGCGCGTGGCGCGAATMAACGCGTGCATCATATAT 726
 6632 laalaleu...Proval...AlaphemethrAlatrp..... 6642
 727 CAAGATGAATTAACATTCGCGCTTGTTCAGACGCGCTGTAACAC 776
 6643 tyrglyleuvalhisleuGllyargleuargAla.....G1 6654
 777 CGAGCGCGTGAAT.....GCCCTAGGTGGTTCAGATCAACAAC 817
 6654 ygluarlyleuileuhsleuAlatnrglylythrglyleuAlaAlav 6671
 818 CGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
 6671 alginilealarghts...leuGllyalagluilephealatrAlagly 6686
 852 852
 6687 ThrprogluylsargalatrleuargluuInglyleuAlaAlaAla 6703
 853TGGCAAT.....ACTCGG 868
 6703 laspserargserleuaspphealagluInvalleuAlaAlaAlaAla 6720

```

869 GCGAATTGGTTGACACAGACACCGCGTGTATTCGGTTCGGTATTGAC 918
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6720 TgtlgllyValaspValValleuAsnSerleuSerlyAlaAlaIleasp 6736
      : : : : : : : : : : : : : : : : : : : : : : : :
919 GGGCGCATTAACACAGCGCGACGATTAATTTGGACGCTACACATCA 968
      : : : : : : : : : : : : : : : : : : : : : : : :
6737 AlAserleuSer..... 6740
      : : : : : : : : : : : : : : : : : : : : : : : :
969 GATTTCGGTTATGACAGAGCGCGACGAGAGCTGTTCGGTGGTGG 1018
      : : : : : : : : : : : : : : : : : : : : : : : :
6741 ....ThreValAlProAspGlyAlaPheIleIleu..... 6751
      : : : : : : : : : : : : : : : : : : : : : : : :
1019 CGCGCAGCGCGACAAATACCTCCATCAGCGCTACAAACCTCGCCATTTC 1068
      : : : : : : : : : : : : : : : : : : : : : : : :
6752 ..GlyLysThrAspIleTyrAlaAspArgSerleuGlyLeuAlaHisPhe 6767
      : : : : : : : : : : : : : : : : : : : : : : : :
1069 CTGAATAACAACTCTTCAGTTCAACACACACCTCAAGCGCGGACCG 1118
      ||||| : : : : : : : : : : : : : : : : : : : :
6768 ArgLysSerleuSerTyr.....SerAlaVal.....AspLe 6778
      : : : : : : : : : : : : : : : : : : : : : : : :
1119 CGCCATGTGTGGCGATTGGTACTTACGAGCGCGT..... 1152
      ||||| : : : : : : : : : : : : : : : : : : : :
6778 uAlaGlyLeuAlaValAlaArgArgProGluArgValAlaAlaLeuAlaG 6795
      : : : : : : : : : : : : : : : : : : : : : : : :
1153 .....ATGCCCTTGGAT 1164
      : : : : : : : : : : : : : : : : : : : : : : : :
6795 LuValAlaAspLeuLeuAlaArgGlyAlaLeuGlnProLeuProValGlu 6811
      : : : : : : : : : : : : : : : : : : : : : : : :
1165 ATCTGCGCCACCTGCTTTTGGCGGATTAATGTCGCGGATACCGACAG 1214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6812 IlePheProLeuSerArgAlaAlaAspAlaPhe..ArgLysMetAlaGln 6827
      : : : : : : : : : : : : : : : : : : : : : : : :
1215 CGCGCAGGCAATGGTGTCTTGGATTTGACAGAGAGACCTCGCTTGT 1264
      : : : : : : : : : : : : : : : : : : : : : : : :
6828 AlaAlaHisLeuGlyLysLeuValLeuAlaLeuGluAspProasp...Va 6843
      : : : : : : : : : : : : : : : : : : : : : : : :
1265 GCACCTTCGTCTGCGCGCGCA 1286
      : : : : : : : : : : : : : : : : : : : : : : : :
6843 ArgIleArgValProGlyGlu 6850
      : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:US-09-568-472-5

```

seq_documentation_block:
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zilkke, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR APPLICATION NUMBER: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5

```

alignment_scores:
Quality: 103.50 Length: 541
Ratio: 0.452 Gaps: 27
Percent Similarity: 42.329 Percent Identity: 21.442

alignment_block:
US-09-303-518d-125 x US-09-568-472-5
Align seg 1/1 to: US-09-568-472-5 from: 1 to: 7257

```

55 GCGGTTTACAGAGCGCGCATTTACGAGCGGCTTGGCGAA.. 102
      ||| : : : : : ||| : : : : : ||||| : : : : :
6354 AlAserPheAspGlyGlnAlaProHisSerValAlaHisLeuGlySerLe 6370
      : : : : : : : : : : : : : : : : : : : : : : : :
103 GAATTCGCGGTATGCCCGCCCTCGATGAAGTCAAGAGAGCGATCCG 151
      : : : : : : : : : : : : : : : : : : : : : : : :
6370 uAspGluArgGlyValLeuAspAlaAspAlaProPheAspAlaSplal 6387
      : : : : : : : : : : : : : : : : : : : : : : : :
152 TCMAAAGAGC.....CAAGTCGTTTGAAGACAA 183
      : : : : : : : : : : : : : : : : : : : : : : : :
6387 euGluGlnSerleuValArgGlyCysAspSerValIleTyrPheValGln 6403
      : : : : : : : : : : : : : : : : : : : : : : : :
184 AAGAATCCGGCGCTGTGTCTTACTGCGCG.....GCTTC 218
      ||||| : : : : : ||||| : : : : :
6404 AlaValAlaGlyAlaGlyPheArgAspProArgLeuTyrPheValIth 6420
      : : : : : : : : : : : : : : : : : : : : : : : :
219 AGGCAAAATCCCGCGGATTCACCGTGGCGAAAGCGGTACTTCACTCAG 268
      : : : : : : : : : : : : : : : : : : : : : : : :
6420 rArgGlyAlaGlnAlaIleGlyAlaGlyAspValSerValAlaGlnAla 6437
      : : : : : : : : : : : : : : : : : : : : : : : :
269 TC.....GTGATTCGCGTGAA..... 285
      : : : : : : : : : : : : : : : : : : : : : : : :
6437 rGluLeuGlyLeuGlyArgValIleAlaLeuGlnHisAlaGluLeuArg 6453
      : : : : : : : : : : : : : : : : : : : : : : : :
286 .....GGCAACGACGAATGAGTTGACG 311
      : : : : : : : : : : : : : : : : : : : : : : : :
6454 CysAlaArgIleAspLeuAspProAlaArgArgAspGlyValAlaSpl 6470
      : : : : : : : : : : : : : : : : : : : : : : : :
312 CTACGACCTGAAGCGCTGCAACTTAAGCGCGCAAGAGCGCGCGCA 361
      : : : : : : : : : : : : : : : : : : : : : : : :
6470 uLeuLeuAlaGluLeuLeuAlaAspAspAlaGluGluValAlaAlaPhe 6487
      : : : : : : : : : : : : : : : : : : : : : : : :
362 ACCGTATCCATCCGGTTGTGTGACTGCGCTCGCACCGCGCTTTCAGC 411
      : : : : : : : : : : : : : : : : : : : : : : : :
6487 rGluGlyGluArgArgValAlaArgLeuValAlaArgLeuProGluIth 6503
      : : : : : : : : : : : : : : : : : : : : : : : :
412 .....AAATTCCTGCGCGATGCGGACCGCGTTCGCATCTT 449
      ||||| : : : : : : : : : : : : : : : : : : : :
6504 AspCysArgGluIleuIleGluProAlaGluGlyArgProPheArgLeuG1 6520
      : : : : : : : : : : : : : : : : : : : : : : : :
450 CGTCAT.....CGATGACCA 466
      : : : : : : : : : : : : : : : : : : : : : : : :
6520 uIleAspGlySerGlyValLeuAspAspLeuValLeuArgAlaIthGln 6537
      : : : : : : : : : : : : : : : : : : : : : : : :
467 CCAATCCGCTGCGCGCGACCGCTACCGCATTAACAAAGAGCGCGAG 516
      : : : : : : : : : : : : : : : : : : : : : : : :
6537 rArgProProGlyProGlyGluValGluIleAlaValGlnAlaAla... 6552
      : : : : : : : : : : : : : : : : : : : : : : : :
517 GATTTCAAACGCGCGCTGTGTGATTTGAGCGTTTGACCGCAACCAAT 566
      ||||| : : : : : : : : : : : : : : : : : : : :
6553 .....GlyLeuAsnPheLeuAspValIleThrArgAlaMetGlyI 6565
      : : : : : : : : : : : : : : : : : : : : : : : :
567 CCAT.....GTTGTAAAG. 579
      : : : : : : : : : : : : : : : : : : : : : : : :
6565 eTyrProGlyProGlyAspGlyProValAlaLeuGlyAlaGluCysSerG 6582
      : : : : : : : : : : : : : : : : : : : : : : : :
580 .....GCAAGTCGCGCAGACGTCGCGT.....GAA 606
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6582 TArgGlyLeuAlaIleMetGlyGluGlyAlaGlnSerleuAlaGlyIleGln 6598
      : : : : : : : : : : : : : : : : : : : : : : : :
607 AATGTCGCAACATCGAACAACATGAATTGCG..... 639
      : : : : : : : : : : : : : : : : : : : : : : : :
6599 AspValValAlaValAlaProPheSerPheGlyThrHisValThrIleAs 6615
      : : : : : : : : : : : : : : : : : : : : : : : :
640 .....GGCCGCATCCTGCGCGGTGTGATGAGCGACGACATTC 676
      : : : : : : : : : : : : : : : : : : : : : : : :
6615 rAlaArgMetLeuAlaProArgProAlaAlaLeuThrAlaAlaGlnAla 6632
      : : : : : : : : : : : : : : : : : : : : : : : :

```

```

677 ATTTCATCGAGCGGTCGGCGCAATAAACCTGTGACCATCAATTAT 726
      ::::: ||||| ||| ||| |||
6632 laalaleu...Proval...Alaphemethralatrp... 6642
      ::::: ||||| ||||| |||||
727 CAAGATGTAATACCATTTGGCTGTGTGCAACAGCGCTGTGAACAC 776
      ::::: ||||| ||||| |||||
6643 TyrglyleuValHisLeuGlyArgLeuArgAla... 6654
      ::::: ||||| ||||| |||||
777 CGAGCGCGTGAT...GCCCTAGGTGGTTCATCAAGTCACAAAC 817
      ||||| ||||| ||||| |||||
6654 ygluArgValleuHisSerAlaThrGlyThrGlyleuAlaLav 6671
      ::::: ||||| ||||| |||||
818 CGCGCGCTTGGCTACCGCTTTGGTGGCGAAAGTA... 852
      ::::: ||||| ||||| |||||
6671 alGlnIleAlaArgHis...LeuGlyAlaGlnIlePheAlaThrAlaGly 6686
      ::::: ||||| ||||| |||||
852 ..... 852
6687 ThrProGlnIuLysArgAlaTrpLeuArgGlnGlnIleAlaHisValMe 6703
853 ..... TCGCAAT...ACTGCG 868
6703 tAspSerArgSerLeuAspPheAlaGlnIleValleuAlaAlaThrLysG 6720
869 GCGAATTTGGTGACACAGACACACCGCGTGAATTCGGTTCGTAATGAC 918
      ||||| ||||| ||||| |||||
6720 IyGlnIleValAlaSpValAlaLeuAsnSerLeuSerGlyAlaAlaIleasp 6736
919 GCGCGCATTAACACAGCGCGCACGATTAATTTGGAGCTTACCAATCA 968
6737 AlaserLeuSer... 6740
969 GATTCGCGTATCGAAGAGCGCGCACAAAGCTGTTCGGCTGGGTTG 1018
6741 .....ThrleuValIProaspGlyArgPheIleGlnleu... 6751
1019 CGCGCGACGCGGACAAATCTCGATCAGCGGTACAAACCTGGGCAATTC 1068
      ::::: ||||| ||||| |||||
6752 ..GlyLysThrAspIleTyrAlaAspArgSerLeuGlyleuAlaHisPhe 6767
1069 CTGAAATAACACTCTCAAGTTCACACAGCGCTCACAGGCGGCGACCG 1118
      ::::: ||||| ||||| |||||
6768 ArgLysSerLeuSerTyr...SerAlaVal...AspLe 6778
1119 CGCGATGTCGCGATGCTGACTTACGAGCGCTG... 1152
      ||||| ||||| ||||| |||||
6778 uAlaGlyleuAlaValArgArgProGlnArgValAlaAlaIleleuAlaG 6795
1153 .....ATGCCCTTGAT 1164
6795 IuValValaSpLeuLeuAlaArgGlyAlaLeuGlnProleuProValGln 6811
1165 ATCTGCGCCACCTGCTTTGGCGGATTAATCGTGGCGATACCGACAG 1214
      ||||| ||||| ||||| |||||
6812 IlePheProleuSerArgAlaAlaAspAlaPhe...ArgLysMetAlaGln 6827
1215 CGCGACGCGATTTGGTGTGTAATTCAGCAAGACCTGCGCTTTGT 1264
6828 AlaGlnHisLeuGlyLysLeuValleuAlaLeuGlnAspProasp...Va 6843
1265 GCAGCTTCGTCGCGCGGCA 1286
      ||||| ||||| ||||| |||||
6843 IArgIleArgValProGlnIu 6850

```

seq_name: /cgn2_6/ptodata/1/iaa/5a_COMB.pep:us-08-046-585-5

seq_documentation_block:

: Sequence 5, Application US/08046585

: Patent No. 5453362

: GENERAL INFORMATION:

: APPLICANT: Lamarco, Kelly

: APPLICANT: Wilson, Angus

: APPLICANT: Herr, Winship

```

? TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
? TITLE OF INVENTION: HOST CELL FACTOR
? NUMBER OF SEQUENCES: 15
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: CA
? COUNTRY: USA
? ZIP: 94111-4187
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent in Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/046,585
? FILING DATE: 12-APR-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Osman, Richard A
? REGISTRATION NUMBER: 36,627
? REFERENCE/DOCKET NUMBER: A-57503-1/RAO
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 781-1989
? TELEFAX: (415) 398-3249
? TELEX: 910 277299
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2035 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-046-585-5

```

```

alignment_scores:
  Quality: 100.00      Length: 472
  Ratio: 0.476         Gaps: 26
  Percent Similarity: 44.492   Percent Identity: 22.034

```

alignment_block:

US-09-303-518D-125/rev x US-08-046-585-5 ..

Align seg 1/1 to: US-08-046-585-5 from: 1 to: 2035

```

1112 CGCGCGTTGACGCGCTGTGTTG.....AACTGAGAGCTTTGTTT 1072
      ||||| ::||| ||||| |||||
366 ProProAlaArgValGlnleuValArgAlaAsnThrAsnSerLeuGlnVa 382
      ::||| ||||| ::::: |||||
1071 CAGGAATGCGCGAGGTGTGACGCGTGATGAGATTTG..... 1032
      ::||| ||||| ::::: |||||
382 L...SerTrpGlyAlaValAlaThrAlaAspSerTyrLeuGlnleuG 398
      ::||| ||||| ::::: |||||
1031 .....TCCGCGTCCGCGCGCAACCCAGCGCAAGAC 1002
      ::||| ||||| ::||| |||||
398 IuLysTyrAspIleProAlaThrAlaAlaThrAlaSerProThr... 413
      ::||| ||||| ::||| |||||
1001 TCTTTGCTGCGGCTTCTTGTGATACGGAATGATGATGATGATGATGATG 952
      ||||| ||||| ||||| |||||
414 .....ProAsnProValProSerValProAla...AsnPr 424
      ||||| ||||| ||||| |||||
951 CAATTAATCGTGGCGCTGTGTATCGCGCGTTC.....ATA 911
      ||||| ||||| ||||| |||||
424 OProLysSerProAlaProAlaAlaAlaAlaProAlaValGlnProleuT 441
910 CCGAACCGGAATGACGCGGTTGTCTGTTCACCAATTCG..... 870
      ||||| ||||| ||||| |||||
441 hGlnValGlyIleThrleuLeuProGlnAlaAlaProAlaPropProThr 457
869 .....CCGCGATAATTTGCATCTTTCGACCC 841
      ||||| ||||| ||||| |||||

```



```

458 ThrThrThrIleGlnValLeuProThrValProGlySerSerIleSerVal 474
840 CAACAGCGGTACGACAGAGCGCGGT.....TTGTGACTTGACAAAC 800
474 LProThrIleAlaIleArgThrGlnGlyValProAlaValLeuLysValIle 491
739 CACCTAGGCGCAATCAGCGCTCG.....GTGTTACAGCGCGCTGT... 759
491 LProGlnAlaIleThrThrGlyThrProLeuValThrMetAlaGlyProAlaSer 507
758 ...GCAACCAACAGCGCAATGTAATACCTTGTAATTAATGATGATCA 712
508 GlnAlaGlyLysAlaProValThrValThrSer.....518
711 CACGGTTTATTTCGCCGCGCAGCGCTGATGAATGATGCGTCCAC 662
519 .....LeuProAlaGlyValArg.....MetValValProT 529
661 TCACACCGCGCA.....GGATGCGGCGCGCGCAATTCATGT 627
529 hrGlnSerAlaGlnGlyThrValIleGlySerSerProGlnMetSerGly 545
626 GTTTCGATGTTGGCAGCATTTTCAGACGCGAG.....594
546 MetAlaIleLeuAlaIleAlaAlaAlaIleThrGlnLysIleProProse 562
593 .TCGCGCGCAGCTGCTTACAA.....ACATGATTTTCGCT 557
562 rSerAlaProThrValLeuSerValProAlaGlyThrThrIleValLysT 579
556 CG.....555
579 hrMetAlaValIleProGlyThrThrThrLeuProAlaThrValLysVal 595
554 .....GTCAACGCGCTACATACCAACAGCGCGCTTGAAATC 517
596 AlaserSerProValMetValSerAsnProAlaIleArgMetLeuLysTh 612
516 CTGGGGGCTCTTGTATATACGCTAGCGGCGCAGCC.....477
612 rAlaAlaAlaGlnValIleGlyThrSerValSerSerAlaThrAsnThrSert 629
476 .....AGCGATGTGTGTCATGCCA... 456
629 hrArgProIleIleThrValIleLysSerGlyThrValThrValAlaGln 645
455 .....TTGACGAGATGCGCAACGCTCGCATCAGC... 423
646 GlnAlaGlnValValThrThrValAlaGlyValThrLysThrIleTh 662
422 .....CGAGATTTTCTGA 407
662 rLeuValLysSerProIleSerValProGlyLysSerAlaLeuIleSerA 679
406 AC...GGACGGGTGCGCAGCGCAGTCACAAACGCGATTGGATCAGGTG 360
679 snLeuGlyLysValMetSerValAlaGln.....688
359 CGGCGCACTCTTCGCGCTTAAGTTCCAGCGCTCAGGTGCGG 310
689 .....ThrLysProValGlnThrSerAlaValThrGlyAlaLase 702
309 TTCAAACTCGATTTCGTCGTG.....288
702 rThrGlyProValThrGlnIleIleGlnThrLysGlyProLeuProAlaG 719
287 .....CCTTACAGCGCA 276
719 LThrIleLeuLysLeuValThrSerAlaAspGlyLysProThrThrIle 735
275 ATCAGAGTACTGAGTACGCGCTTTCGACGCGTGAATCCGCGCAT 226
736 IleThrThr.....738

```

```

225 TTTCGCTGAACCGCGCGCAGTAACACACGCGCGATTCCTTTGCTT 176
729 ...ThrGlnAlaSerGlyAlaGlyThrLysPro...ThrIleLeuGlyI 753
175 CAACAGCACTTGGCTTTTTCAGCGCAGCGCTTCCTGACTTTCATC 126
753 lSerSerValSerPro...SerThrThrLysProGlyThrThrThrIle 768
125 GAGGCGCGCATACCG.....GCATATCTTCCGCAACGACGCGACTTC 82
769 lIleThrThrIleProMetSerAlaIleIleThrGlnAlaGlyValThrG 785
81 GGTATGCGCGCGCGCG 66
785 YValThrSerSerPro 790

```

seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-393-703-5

```

seq_documentation_block:
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Iamarco, Kelly
; APPLICANT: Wilson, Angus
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,703
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SFO ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-393-703-5

```

alignment_scores: Quality: 100.00 Length: 472
Ratio: 0.476 Gaps: 26
Percent Similarity: 44.492 Percent Identity: 22.034

alignment_block:
US-09-303-518d-125/rev x US-08-393-703-5 ..
Align seg 1/1 to: US-08-393-703-5 from: 1 to: 2035

```

1112 CCGCCGTGACGGCTGTGTTG.....ACTTGAGAGTTGTTT 1072
1113 ||||| :||| ||| :||| |||
366 PROProAlaArgValGlnLeuValAlaGalaSerThrIleu 382
1071 CAGGAATGCGCGAGGTTGTACGGCTGATGAGATTTG..... 1032
382 1...SettPglValValAlaThrAlaAspSerThrLeuGlnLeu 398
1031 .....TCCGGCTGGGGCAACCGCGGCAACG 1002
398 InuStrAspIleProAlaThrAlaAlaThrSerProThr... 413
1001 TCCTTGTGGCGCTTCTTCGATAACGAAATCTGTTGTGTCAGCTCC 952
414 .....ProAsnProValProSerValProAla.....AspPr 424
951 CAATAATCTGTGGCGCTGTGTAATCGCGCTTC.....ATA 911
424 OProlYSerProAlaProAlaAlaAlaProAlaValGlnProLeu 441
910 CCGAACCGGAATCAGCGGCTGTGTGTCACCAATTCG..... 870
441 hrgInValGlyIleThrLeuLeuProGlnAlaAlaProAlaProThr 457
869 .....CCCGCAGTAATTGGCATCTTTCGACC 841
458 ThrThrThrIleGlnValLeuProThrValProGlySerSerIleSerVa 474
840 CAATACGCTACGCAAGAGCGCGT.....TGCTTGACTTGGAAC 800
474 1ProThrAlaAlaArgThrGlnGlyValProAlaValLeuLysValThG 491
799 CACCTAGGCAATCAGCGCTCG.....GTGTCAGACGCGCTGT... 759
491 1ProGlnAlaThrThrGlyThrProLeuValThrMetArgProAlaSer 507
758 ...GCAACCAACGCGCAATGTAATTCATCTTGATTAATTGATGCCA 712
508 GlnAlaGlyLysAlaProValThrValThrSer..... 518
711 CACGGTTTATTCGCGCGACCGCTGATGAATGAATGCGTGCAC 662
519 .....LeuProAlaGlyValArg.....MetValValPro 529
661 TCAACCGGCA.....GGATGCGGCGCGCGCAATTCATGT 627
529 hrgInSerAlaGlnIleThrValIleGlySerSerProGlnMetSerGly 545
626 GTTTCGATGTTGGCAGCATTTTCAGACGGCAG..... 594
546 MetAlaAlaLeuAlaAlaAlaAlaAlaThrGlnLysIleProProSe 562
593 .TCTGGCCAGCTGCGCTTACAA.....ACATGATTTGGCTT 557
562 rSerAlaProThrValLeuSerValProAlaGlyThrThrIleValLys 579
556 CG..... 555
579 hrmEtAlaValThrProGlyThrThrThrLeuProAlaThrValLysVal 595
554 .....GTCAACGCTCAATACCAACAGCGCGCTTGAATC 517
596 AlaSerSerProValMetValSerAsnProAlaThrArgMetLeuLysTh 612
516 CTGCGCGGCTTCTTGATTAATACCGTAGGCGCGCAGC..... 477
612 rAlaAlaAlaGlnValGlyThrSerValSerSerAlaThrAsnThrSer 629
476 .....AGCGATTGCTGCATCGCA... 456
629 hArgProIleIleThrValIleLysSerGlyThrValThrValAlaGln 645
455 .....TTGACGAAGATGGCGAAGCGCTCGCATCGACG.... 423

```

```

646 GlnAlaGlnValValThrThrValValGlyValAlaThrThrIleTh 662
422 .....CGAAGATTTTGGTGA 407
662 rleuValLysSerProIleSerValProGlyIleSerAlaLeuIleSer 679
406 AC...GGACGGGTGGCAGCGCAGTCACCAACCGGATGTGATCAGGTTG 360
679 snLeuGlyLysValMetSerValValGln..... 688
359 CCGCGCACTTCTCCCGCTTGAATTTCACGCGCTTCAGGTGAGTACG 310
689 .....ThrLysProValGlnThrSerAlaValThrGlyGlnAlaSe 702
309 TTCAACTGCAATTCGTCGTTG..... 288
702 rThrGlyProValThrGlnIleIleGlnThrLysGlyProLeuProAlaG 719
287 .....CCTTCAACGCGCA 276
719 1yThrIleLeuLysLeuValThrSerAlaAspGlyLysProThrThrIle 735
275 ATCAGCACTGACTGAAGTACGGCGCTTTCGACGAGTGAATCGCGCAT 226
736 1leThrThr..... 738
225 TTGCGCTGAAGCGCGCAGTAAACCAACCGCGGATTCCTTTGTCTT 176
739 .....ThrGlnAlaSerGlyAlaGlyThrLysPro...ThrIleLeuGlyI 753
175 CAACACCACTTGCGCTTTTTCAGCGCAATCGCTTCCTTGAATTCATC 126
753 1eSerSerValSerPro...SetThrThrLysProIleThrThrThrIle 768
125 GAGGGGGCGCATACG.....GCATATCTTTCGCCCAAGCAGCGCATC 82
769 1leLysThrIleProMetSerAlaIleIleThrGlnAlaGlyAlaThrG 785
81 GGTAAATGGCGGCGCG 66
785 yValThrSerSerPro 790

seq_name: /cgn2_6/plodata/1/1aa/pctvus_comb.pep:pct-US93-11721-5
seq_documentation_block:
; Sequence 5, Application PC/TUS9311721
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11721
; FILING DATE: 03-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627

```

REFERENCE/DOCKET NUMBER: PD-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-11721-5

alignment_scores:

Quality: 100.00 Length: 472
Ratio: 0.476 Gaps: 26
Percent Similarity: 44.492 Percent Identity: 22.034

alignment_block:

US-09-303-518D-125/rev x PCT-US93-11721-5

Align seg 1/1 to: PCT-US93-11721-5 from: 1 to: 2035

```
1112 CCGCGGTGACGCGCTGCTTG.....ACTTGAAGAGTTTGTCTT 1072
|||||
366 ProProAlaArgValGlnLeuValArgAlaAsnThrAsnSerLeuGluVal 382
1071 CAGGAATGCGCGGAGGTTGACGCGTATGAGTATG..... 1032
|||||
382 L...SerTrpGlyAlaValAlaThrAlaAspSerTyrLeuGlnLeuG 398
1031 .....TCCGGCTGCGCGCGCAACCGCGCGACAGC 1002
398 LulYrAspIleProAlaThrAlaAlaThrSerProThr... 413
1001 TCTTGTGCGCGCTTCTTGTGATACGGAATCTGATGTGAGCGTCC 952
|||||
414 .....ProAsnProValProSerValProAla...AsnPr 424
951 CAATAATGCTGCGCGCTTGTGTAATCGCGCGTTC.....ATA 911
|||||
424 oProLysSerProAlaProAlaAlaAlaProAlaValGlnProLeuT 441
910 CCGAACCGGAATCAGCGGTTGTCTGTCAACCAATTCG..... 870
|||||
441 hrGlnValGlyIleThrLeuLeuProGlnAlaAlaProAlaProThr 457
869 .....CCCGCAGTAATTTGCGATACTTTCGACG 841
458 ThrThrThrIleGlnValLeuProThrValProGlySerSerIleSerVa 474
840 CAACAGCGTACGAGAGCGCGGT.....TTGTTGACTTGAGAAC 800
|||||
474 lProThrAlaAlaArgThrGlnGlyValProAlaValLeuLysValThrG 491
799 CACCTAGGCAATCAGCGCTCG.....GTGTCAGACGCGCTGTT... 759
|||||
491 lProGlnAlaThrThrGlyThrProLeuValThrMetArgProLysSer 507
758 .....GCAACCAACGCGCAATGTAATCACTTGATTAATGATGATGCA 712
508 GlnAlaGlyLysAlaProValThrValThrSer..... 518
711 CACGGTTTATTCGCGCGCGCGCGTATGAATGAATGATGCTGCGCAC 662
|||||
519 .....LeuProAlaGlyValArg.....MetValValProT 529
661 TCAACCGCGCA.....GATGCGCGCGCGCGAATTCATGT 627
529 hrGlnSerAlaGlnGlyThrValIleGlySerSerProGlnMetSerGly 545
```

```
626 GTTTCGATGTGCGACGATTTTCAACGCGCAG..... 594
546 MetAlaAlaLeuAlaAlaAlaAlaAlaAlaThrGlnLysIleProProSe 562
593 .TCTGCGCGAGCTGCTTACAA.....ACATGATTTTGGCTT 557
|||||
562 rSerAlaProThrValIleSerValProAlaGlyThrThrIleValLysT 579
556 CG..... 555
579 hrMetAlaValThrProGlyThrThrThrLeuProAlaThrValLysVal 595
554 .....GTCAACGCGCTCAATTCACACAGCGCGCTTGAATC 517
|||||
596 AlaSerSerProValMetValSerAsnProAlaThrArgMetLeuLysThr 612
516 CTCGCGCGCTTCTTGTATATGACCGTATGAGTGCAGCGC..... 477
|||||
612 rAlaAlaAlaGlnValGlyThrSerValSerSerAlaThrAsnThrSerT 629
476 .....AGCGATTTGCTTCATCGCA... 456
629 hrArgProIleIleThrValHisLysSerGlyThrValThrValAlaGln 645
455 .....TTGACGAAGATGCGCGAGCGCTCGCGCATCGAG... 423
646 GlnAlaGlnValThrThrThrValGlyGlyValThrLysThrIleThr 662
422 .....CGAGGAATTTGCTCA 407
662 rLeuValLysSerProIleSerValProGlyLysSerAlaLeuIleSerA 679
406 AC...GGACGCGTGCAGCGCGACGCGCACCAACCGGATGATGATGAGTGT 360
|||||
679 snLeuGlyLysValMetSerValValGln..... 688
359 CCGGCGCATCTTTCGCGCTTAACTTGCACGCGCTTCAAGTGCATGACG 310
689 .....ThrLysProValGlnThrSerAlaValThrGlyGlnAlaSe 702
309 TTCAACTGATTCGCTGTTG..... 288
702 rThrGlyProValThrGlnIleIleGlnThrLysLysProLeuProAlaG 719
287 .....CCTTCAACGCGCA 276
719 lYThrIleLysLeuValThrSerAlaAspGlyLysProThrThrIle 735
275 ATCAGCACTGACTGAAGTACGCGCTTTCGCGAGGTAAATCGCGCGCAT 226
736 lIleThrThr..... 738
225 TTTTGCCTGAGCGCGCGCGCAATACACAGCGCGGATTTCTTTTGTCTT 176
739 .....ThrGlnAlaSerGlyAlaGlyThrLysPro...ThrIleLeuGly 753
175 CAACAGCACTTGGCTTTTTCAGCGGATGCGCTTCTTCACTTATC 126
|||||
753 lSerSerValSerPro...SerThrThrLysProGlyThrThrThrIle 768
125 GAGGCGCGCATTCG.....GCATATTCCTGCGCAACGAGCAAGCACTTC 82
769 lIleLysThrIleProMetSerAlaIleIleThrGlnAlaGlyAlaThrG 785
81 GGTATGCGCGCGCGCG 66
785 yValThrSerSerPro 790
```

seq_name: /cgn2.6/ptoda/1/1aa/5B.COMB.pep:us-08-276-967-2
seq_documentation_block:
; Sequence 2, Application us/08276967
; Patent No. 5851817

GENERAL INFORMATION:
 APPLICANT: Hardy, Daniel M.
 TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
 TITLE OF INVENTION: Sperm
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P. O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: USA
 ZIP: 77210-4433
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-08/276,967
 FILING DATE: Submitted Herewith
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: UTSD:418\KIT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 713-787-1400
 TELEFAX: 713-789-2679
 TELEX: 79-0924
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2476 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-276-967-2

alignment_scores:
 Quality: 99.00 Length: 373
 Ratio: 0.615 Gaps: 19
 Percent Similarity: 43.164 Percent Identity: 20.912

alignment_block:
 US-09-303-518D-125 x US-08-276-967-2 ..
 Align seg 1/1 to: US-08-276-967-2 from: 1 to: 2476

```

381 GTGACGTGGCGTGGCGAC..... 398
1849 LeuAspCysSerAlaHisSerValThrSerCysValProSerCysLeu 1865
399 ..CCGTCGCGTTCAG.....CAAAATTCCTGCCGTGCATGCCGAGCC.... 437
1865 uProSerCysGlnAspProGlnGlyGlnCysThrGlyAlaGlyAlaPro 1882
438 .....GTTCGGCACTT 449
1882 eThrCysGlnGlyGlyCysIleCysGluProGlyTyrValLeuSerGlu 1898
450 CGTCATGTGGAGGACACCAATCCGCTGGCTGCCGA..... 485
1899 GlnGlnCysValAlaArgSerGlnCysGlyCysArgAspAlaArgGlyTh 1915
486 .....CCCTACGTCATTATCAAGAAGCCGCCGAGGATTTCAA.... 524
1915 rPheLeuProValGlyArgPheArgLeuSerSerGlyCysSerGlnMetC 1932
525 .....ACGGCGCGCTGTGATTCGACCGTTGACCGAAGCGAA 563
1932 ySValCysThrAlaGlyAlaIleGlnCysArgProPheThrCysProSe 1948

```

```

564 AATCATGTTTGTAG.....GCAGCTGGCGGACGTCGCTGTGAAA 607
1948 rGlySerGlnCysGluProAsnGlnAspGlyLysAspPheCysGlnPro 1965
608 ATGCTGCCACATCGAACAACATGATTCGGCGCCGATTCCTGCCGGT 637
1965 snSerSerAsnLeu...CysSerValPheGlyAspProHisTyrArgThr 1980
658 TTGAGTGGCGGCGACATTTCATTC..... 681
1981 PheAspGlyLeuSerTyrArgPheGlnGlyArgMetThrTyrThrLeu 1997
682 .....ATGACCGCGTC.....G 694
1997 llyThrLeuAspValLeuProAspGlyValGlnProLeuValAlaGlu 2014
695 GCGCGCAATAA.....ACCGTGGACCATCATTCATTCAGATGTA 735
2014 lYArgAsnLysValTyrProSerLeuThrProValPheLeuGlnGlu 2030
736 ATTACCATTCGGCGCTTGTTCACACAGCGCGTGAACCGCGCGT 785
2031 lIleValMetValTyrGlyTyr...ThrValGlnLeuGlnAlaGluLeu 2046
786 GATTGCCCTAGGTGTTCTCAAGTCACAAACCG.....C 820
2046 uLeuValAlaAsnGlyGlnLysValSerIleProTyrLysProAsnGlu 2063
821 GCCTCTTGCGTACCGCTTTGGTGCGCAAGTATCCAAATTCCTCGGCG 870
2063 yrLeuGlnValThrLeuArgGlyArgArgLeuTyrLeuValThrAsp 2079
871 GAATGTGTGACACAGACAAACCGGATTCGCTCGTATTCAGCG 920
2080 GluLeu.....ValValSerPheAsnGlyArgAsn 2090
921 CGCGATTACACAAAGCGCGCATTTATTTGGAGCGCTACCAATACAGA 970
2090 nAlaValIleAlaMetProSerThrTyrLeuGly..... 2101
971 TTTCGTTATCGAAGAAGCGCGCAGCAAGAGCTGTTCGCTGGTTCG 1020
2101 ..... 2101
1021 CCGAGCGCGAACAATCTCCATCAGCGCTACCAACCTGGCGCATTCCT 1070
2102 .....LeuValArgGlyLeuCysGlyAsnTyrAs 2111
1071 GAAACAAACTCTTCAAGTTCAACACAGCGCGTCAACGCGCGGACCGG 1120
2111 pLysAsnLysArgAsnAspPheMetLeuProAsnGlySerPheThr 2127
1121 CCAT.....GGTCCGATTGGTACTTACAGCGCGCGATGCC 1158
2127 nasnLeuValPheGlyAsnSerTyrGluValLysGlu.... 2142
1159 TTGGATATCCGCGCCACCGCTTTCGGCGATTCATTCGCGCGATAC 1208
2143 .....GlyHisProArgPheSerArgAlaIleArgGlnGlu 2155
1209 CGACAGCGCGAGCGCATTCGCTGGATTCGAAATTCGACGAGAACACTCG 1258
2156 GluLysAsnGlnGluSerGlyPheGlnAsnValSerGlyCysSerPro 2172
1259 CTTTGTGCAGCTT 1271
2172 uGlnLeuGlnLeu 2176

```

seq_name: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:US-08-997-897-2
 seq_documentation_block:
 ; Sequence 2, Application US/08997897C

```

1 Patent No. 6114514
2 GENERAL INFORMATION:
3 APPLICANT: SRIVASTAVA, RANJANA
4 APPLICANT: KUMAR, DEEPAK
5 APPLICANT: SRIVASTAVA, BRAHM SHANKER
6 TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
7 FILE REFERENCE: u011466-7
8 CURRENT APPLICATION NUMBER: US/08/997,897C
9 NUMBER OF SEQ ID NOS: 7
10 SOFTWARE: PatentIn Ver. 2.0
11 SEQ ID NO 2
12 LENGTH: 430
13 TYPE: PR1
14 ORGANISM: Mycobacterium tuberculosis
15 FEATURE:
16 NAME/KEY: UNSURE
17 LOCATION: (4)
18 FEATURE:
19 NAME/KEY: UNSURE
20 LOCATION: (6)
21 FEATURE:
22 NAME/KEY: UNSURE
23 LOCATION: (20)
24 FEATURE:
25 NAME/KEY: UNSURE
26 LOCATION: (29)
27 FEATURE:
28 NAME/KEY: UNSURE
29 LOCATION: (54)
30 FEATURE:
31 NAME/KEY: UNSURE
32 LOCATION: (64)
33 FEATURE:
34 NAME/KEY: UNSURE
35 LOCATION: (69)
36 FEATURE:
37 NAME/KEY: UNSURE
38 LOCATION: (89)
39 FEATURE:
40 NAME/KEY: UNSURE
41 LOCATION: (99)
42 FEATURE:
43 NAME/KEY: UNSURE
44 LOCATION: (114)
45 FEATURE:
46 NAME/KEY: UNSURE
47 LOCATION: (119)
48 FEATURE:
49 NAME/KEY: UNSURE
50 LOCATION: (129)
51 FEATURE:
52 NAME/KEY: UNSURE
53 LOCATION: (159)
54 FEATURE:
55 NAME/KEY: UNSURE
56 LOCATION: (169)
57 FEATURE:
58 NAME/KEY: UNSURE
59 LOCATION: (182)
60 FEATURE:
61 NAME/KEY: UNSURE
62 LOCATION: (185)
63 FEATURE:
64 NAME/KEY: UNSURE
65 LOCATION: (219)
66 FEATURE:
67 NAME/KEY: UNSURE
68 LOCATION: (259)
69 FEATURE:
70 NAME/KEY: UNSURE
71 LOCATION: (269)
72 FEATURE:

```

1	NAME/KEY: UNSURE
2	LOCATION: (291)
3	FEATURE:
4	NAME/KEY: UNSURE
5	LOCATION: (323)
6	FEATURE:
7	NAME/KEY: UNSURE
8	LOCATION: (339)
9	FEATURE:
10	NAME/KEY: UNSURE
11	LOCATION: (349)
12	FEATURE:
13	NAME/KEY: UNSURE
14	LOCATION: (356)
15	FEATURE:
16	NAME/KEY: UNSURE
17	LOCATION: (366)
18	FEATURE:
19	NAME/KEY: UNSURE
20	LOCATION: (410)
21	FEATURE:
22	NAME/KEY: UNSURE
23	LOCATION: (421)

```
alignment_scores:
  Quality: 98.00
  Ratio: 0.838
  Percent Similarity: 36.909
  Length: 317
  Gaps: 16
  Percent Identity: 22.082
```

```
alignment_block:
```

US-09-303-518D-125 x US-08-997-897-2

Align seg 1/1 to: US-08-997-897-2 from: 1 to: 430

[illegible]

```

281 GlycysArgGlyCysArgCysGlnGlyCys***SerHisArgArgPheAr 297
482 ..... 482
297 GcysArgGlyCysCysAsnGlyArgCysTrpArgCysArgGlyCysSerA 314
483 .....CGACCCCTACGGTCATTATCA 503
314 rGArgProGlyLeuProGlyArgAsp***ArgProValGlyHisArgGlyS 330
504 .....AGAACCCCGAGAGATTTCAAACGGCGCT. 533
331 IleProThcCysCysPheArgCys***ArgSerProArgSerArgProAl 347
534 .GTTGGTATTGACCGCTTGACCGAAGCAAAATCCATGTTTGTAAAGCA 582
347 aLeu***TrpProGlySerCys***ThrAsnProIle..... 360
583 GCTGGCGCAGACGTGCCGTCTGAAATGCTGCCAACATCGAAACACATGA 632
361 .....ArgCysCysProSer***Ser..... 367
633 ATTCGGCGCCGCATCTGCCG.....TTTGA 661
368 .....ArgProIleProAlaArgProAlaArgLeuProGlyArgSerTrpAr 382
662 GTGG.....CACGCACATTCATTTCATCGA...G 687
382 gTrpProProThrlYsSerGlyArgSerGlnAsnCysTrpHisArgSerS 399
688 CCGGTCCGCGCGAATAAACCGTGTGGACCATTAATTAACAAGA..... 731
399 eArgYserArg.....ThrArgTrpArgHisArgCys***ArgGlyPro 413
732 TGTATATACATTTGGCGCTTGTGTCACACAGCGCTGACACACGAGC 781
414 ThrThrAlaHisTrpSerSer***ArgSerArgAlaProGlyAsnGlnAl 430
782 G 782
430 a 430
seq_name: /cgn2_6/ptoddata/1/1aa/6B_COMB.pep:us-09-370-368-5
seq_documentation_block:
; Sequence 5, Application US/09370368
; Patent No. 6258932
; GENERAL INFORMATION:
; APPLICANT: Anders Vahline
; TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.003A
; CURRENT APPLICATION NUMBER: US/09/370,368
; CURRENT FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Simian Immunodeficiency Virus
US-09-370-368-5

```

```

alignment_scores:
  Quality: 95.50      Length: 322
  Ratio: 0.632      Gaps: 24
  Percent Similarity: 46.894      Percent Identity: 23.913

```

```

alignment_block:
US-09-303-518d-125 x US-09-370-368-5 ..

```

```

Align seg 1/1 to: US-09-370-368-5 from: 1 to: 510

```

```

516 GGATTTCAAACCGCGCTGTGTGATTGAGCCGTTGACCGACGCAAA 565
|||||..... 565
173 GlyPheGlnAlaLeuSerGlnGlyCysThrProTrpHisLeuSln.M 189
|||||..... 189
566 TCCATGTTGT.....AAGCAGCTGCG.....GCAGAC 594
|||||..... 594
189 eLeuAsnCysValGlyGlnHisGlnAlaMetGlnIleLeuArgGlu 205
595 GTCCGCTGAAATGCTGCCACATCGAACA...CATGATTCGGCGG 641
|||||..... 641
206 IleLeuSlnGlnGlnAlaAlaAspTrpAspValGlnHisProGlnProG 222
642 CCCGATCCTCGCGGCTTG.....AGTGCACGACATTCATT 679
|||||..... 679
222 yProLeuProAlaGlyGlnLeuArgGlnProSerGlySerAspIle.... 237
680 TCATCGACCGCGTGGCGGAATAAACCGTG.....TGG 714
|||||..... 714
238 .....AlaGlyThrThrSerThrValGlnGlnGlnIleGlnTrp 250
715 ACCATCAATTACAGATGATATTCATGCGCGTGTGTCACAGG 764
|||||..... 764
251 MetHisArgGlnGlnAsnProIleProValGlyAsnIleTrp..... 264
765 CCGTTCGACACCGACGCGCGTATTGCCCTAGCGTGTCTCAAGTCACA 814
|||||..... 814
265 .....ArgArgTrpIleGlnLeuGlyLeuGlnLysCysVal 276
815 AACCGCGCTCTTGGCTA..... 832
277 ArgMetLysTrpProAlaValAsnIleLeuAspIleLysGlnGlyProLysG 293
833 .CGGTT.....TGGTGCGAAGATGCAATTAATTCGCGGCG... 871
|||||..... 871
293 uProPheGlnSerTrpValAspArgPheTrpLysSerLeuArgAlaGlnG 310
872 .....AATGCTGACACAGACAAACCGCGATTT 901
|||||..... 901
310 LnaIaAspProAlaValLysAsnTrpMetThrGlnThr..... 922
902 CCGGTGCTGATTGAAAGCGCGCGATTCACAAAGCGCGACGATTTTGG 951
|||||..... 951
323 ProLeu.....IleGlnAsnAlaAsnPro..... 330
952 GGACGCTACCAATACATGATTTCCGTTATCGAAGAAGCGCGACGACAAGA 1001
|||||..... 1001
331 AspCysLysLeuValLeuLysGlyLeuGlyMetAsnProThrLeuGln 346
1002 GCTGTTCGGCTGGGTGCGCGACGCGA..... 1031
|||||..... 1031
347 GluMetLeuThrAlaCysGlnGlyValGlyLysProGlyGlnLysAlaAr 363
1032 .....CAATTAATCATCAGCGGTACAAACCTCGCCATTTCCT. 1070
|||||..... 1070
363 gLeuMetAlaGlnAlaLeuLysGlnAlaPheGlnProGlyProLeuProP 380
|||||..... 380
1071 .....GAAACAAACCTGTCAGATTCACACAC... 1097
|||||..... 1097
380 heAlaAlaAlaGlnGlnGlnGlyArgArgTrpValLysCysTrpAsnGly 396
1098 .....AGCCGTCAACGGCGCGACCGCGCATGCTGC 1129
|||||..... 1129
397 GlyLysGlnGlyHisThrAlaLysGlnLysLysAlaProArgArgGlnG 413
|||||..... 413
413 yCysTrp.....LysCysGlyLysProGlyHis...G 423
|||||..... 423
1180 CTTTGGCGGATTTAATCGTCGCGATACCGACGCGCGACGAGCATTTGG 1229
|||||..... 1229
423 lMetAlaLysCysProGlnArgGln.....ValGly 433
|||||..... 433

```



```

890 AC.....CGCGTATTCCGGTTCGGTATTGACGGCGCAT 927
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 snheargilelysarqylleerlyserlythrlleasnpoglylle 415
892 ACACAGAGCGCGCAGCATTTATTTGGACCGCTACCAATCAGATTTCGGT 977
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 ALHhISGlyleaserasptyrvalglyser..... 425
978 TATCGAAGAGCGCGCAGCAAGAGCTGTCCGCTGG..... 1014
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 ValGIuvalIGlylsTyrAlaaspLeuValLeuTrpSerProAlaPheP 442
1014 ..... 1014
442 heGIlylelyProAsnMetIleIlelySGlylyPheIleAlaLeuSer 458
1015 .....GTTCGCGCGCGCAGCGCGCAAAATA 1037
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
459 GluMetGlyAspAlaAsnAlaSerIleProThrProGlnProValTyrTy 475
1038 CTCATACCGCGTACAAACCTCGGCATTTCCGAAAAACAACCTTCA 1087
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 r.....ArgGluMetPheGlyHisGlyHisGlyLysAsn.....L 486
1088 AGTTCACACAGACCGCTCAC..... 1107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 yspheasptHrAsnIleThrPheValSerGlnAlaIaIaTyrLysAlaGly 502
1108 .....GGCGCGACCGCGCATGTGCGCATTCGTACTTA 1142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
503 IlelySGluGluLeuGlyLeuAspArgAlaIaIaProValLysAsnGly 519
1143 CGAGCGCGGTGATGCCCTTGATATC 1167
519 sarGAsnIleThrLysLysAspLeu 527
seq_name: /cgn2.6/prodata/1/laa/5B.COMB.pep:US-08-432-697-21
seq_documentation block:
Sequence 21. Application US/08432697
Patent No. 6248330
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESS: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-00000

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: peptide
TOPOLOGY: linear
FEATURE:
NAME/KEY: Protein
LOCATION: 1..569
OTHER INFORMATION: /note= "URE B - FIGURE 3."
US-08-432-697-21

alignment_scores:
Quality: 95.00 Length: 425
Ratio: 0.562 Gaps: 21
Percent Similarity: 39.765 Percent Identity: 20.706

alignment_block:
US-09-303-518d-125 x US-08-432-697-21 ..
Align seg 1/1 to: US-08-432-697-21 from: 1 to: 569

205 ACTGCGCGCGCTTCAGCAAAATCGCGCGCATTCACCGTGGCAAAACG 254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 ThrGlyProAlaAspGlyThrAsnAlaIaIaThrIleThrProGlyArgAl 177
255 CGTACTCAGTCAGTCGATTCGCGCTTGAAGGCAACGCAAAATCAGT 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177 AsnIleLysSerMetLeuArgAlaIaIa..... 186
305 TTGAACGCTACGCACTGAA.....GCGTGGCAAACTTAAGC 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 ..GluGluTyrAlaMetAsnLeuGlyPheLeuAlaLysGlyAsnValSer 202
343 GCGCAAGAGTGGCGCGCATGATCGCAATCGGTTG..... 381
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 TyrGluProSerLeuArgAspGlnIleGluAlaGlyAlaIleGlyPheLy 219
382 .....TGACTGCGCTGCGCACCGCTCCGTTACGAAATTC 418
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 sIleHisGluAspThr..... 224
419 CTGCGCTGATGCGGACCGCTTGCCTTC.....GTCAATGGATG 462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 .....GlySerThrProAlaIleHisIshiscysLeuAsnValAla 238
463 GACACCAATCCGCTG...GTCGCGACCTACGCTCATTTATCAAGAGC 509
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 AspGluTyrAspValGlnValAlaIleHisThrAspThrLeuAsnGluI 255
510 C.....GCCGAGATTTCAAACCGCGCTGTGTTAGCGCGTTGA 553
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
255 aGlyCysValGluAsp.....ThrLeuGluAlaIleA 266
554 CCGAAGCAAAATCCAGTGTGTAAGCAGCTGCGCATGCCCTCT 603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 IagIyArgThrIleHisThrPheHisThrGluGlyAla..... 278
604 GAAATGCTGCCAATCAGAAACATGAAATTCGCGCGCGCATCTGC 653
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
279 .....GlyGlyGlyHisAlaIaP 284
654 CGTTTGAAGTGCACGCAATTCATTTG...ATGAGCGGCTGGCGCGCA 700
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 AspValIleLysMetAlaGlyGluPheAsnIleLeuProAlaSerThra 301
701 ATAAACCGTG...TGACATCAAT.....TATCAAGATGTA 735
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

301 snProthrlleProPheThrLysasnThrGluAlaGluHisMetAspMet 317
726 ATTACATTGGCCGT.....TTGTT 755
318 LeuMetValLysHisHisLeuAspLysSerLleLysGluAspValGlnPh 334
756 TGCAACAGCGCGTCTGAACACGACGCGGATTCGCCGATTCGCGTTC 805
334 eAlaAspSerArgLleArgProGlnThrIleAla.....GluAspG 349
806 AAGTCACAAACCGCGCTCTTCGTAACGCTTTGGTGCGAAAGTATCG 855
349 InLeuHisAspMetGlyLlePheSerLleThrSerSerAspSerGlnAla 365
856 CAATTCAGTGGCGCAATGGTT..... 879
366 MetGlyArgValGlyGluValIleThrArgThrTrpGlnThrAlaAsp 382
880 .....GACACAGACA 889
382 sAsnLysLysGluPheGlyArgLleLysGluLysGluLysGluAspAsnAsp 399
890 AC.....CCGTGATTCGGGTGGTATTCGAACGCGCGAT 927
399 snPheArgLleLysArgTyrLleSerLysTyrThrLleAsnProGlyLle 415
928 ACACAGGCGCGCGATTCATTTGGACGCTACCAATCAGATTCGCT 977
416 AlaHisGlyLleSerAspTyrValGlySer..... 425
978 TATCAAGAAGAGCGCGACGAAGAGCTGTTCGGCTG..... 1014
426 ValGluValGlyLysTyrAlaAspLeuValLeuThrPseProAlaPhe 442
1014 ..... 1014
442 heGlyLleLysProAsnMetLleLysGlyLysPheLleAlaLseuSer 458
1015 .....GTGGCGCGCGCGGACGAATA 1037
459 GlnMetGlyAspAlaAsnAlaSerLleProThrProGlnProValTyrT 475
1038 CTCATCAGCGGTACACCTCGGCGCATTCCTGAAACAACTCTTCA 1087
475 r.....ArgGlnMetPheGlyHisHisGlyLysasn.....L 486
1088 ACTTCACACAGCGCTCAAC..... 1107
486 ySPheAspThrAsnLleThrPheValSerGlnAlaAlaTyrLysAlaGly 502
1108 .....GGCGGCGACGCGCGCATTCGTGCGGANTGGTACTTA 1142
503 IleLysGluGluLysGlyLeuAspArgAlaAlaProProValLysAsn 519
1143 CGAGCGCGTATGCCCTTGATATC 1167
519 sArgAsnLleThrLysLysAspLeu 527

seq_name: /cgn2_6/prodata/1/1aa/6B_COMB.pep:us-08-466-248-21
seq_documentation_block:
; Sequence 21, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Faradow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..569
; OTHER INFORMATION: /note= "URE B - FIGURE 3."
; US-08-466-248-21

alignment_scores:
Quality: 95.00 Length: 425
Ratio: 0.562 Gaps: 21
Percent Similarity: 39.765 Percent Identity: 20.706

alignment_block:
US-09-303-518D-125 x US-08-466-248-21 ..
Align seg 1/1 to: US-08-466-248-21 from: 1 to: 569

205 ACTGCGCGCGGCTTCAGCAAAATCGCGCGATTCACCGTGGCAAAAGCG 254
||||:|||||.....
161 ThrGlyProAlaAspGlyThrAsnAlaThrThrLleThrProGlyArgAl 177
255 CGTACTTCAGTCAGTCGTGATTCGCCGTGAAGCAACGCAAAATCGAGT 304
||||:|||||.....
177 aAsnLeuLysSerMetLeuArgAlaLa..... 186
305 TTGAACCTAGCGACACTGA.....GCGGTGGCAAACTTAACG 342
||||:|||||.....
187 ..GluGluTyrAlaMetAsnLeuGlyPheLeuAlaLysGlyAsnValSer 202
343 GCGCAAGAGTGGCGCGCATTCGATCAATCGGTTG..... 381
203 TyrGluProSerLeuArgAspGlnIleGluAlaGlyAlaIleGlyPheL 219
382 .....TGACATCGCGCTGGCGCACCGCGTCCGTTGACAAAAATTC 418

```

```

219 silehISgluAspTrp..... 224
219 silehISgluAspTrp..... 224
419 GTCCCGTCGATGCCGAGCCGTCGCATGTC.....GTCAATCCGATG 462
225 .....GlySerThrProAlaAlaIleHisHisCysLeuAsnValAla 238
463 GACACCAATCCGCTG...GTCGCCGACCTACGGTCATATATCAAGAAGC 509
239 AspGluTyrAspValGlnValAlaIleHisThrAspThrLeuAsnGluAl 255
510 C.....GCCGAGATTTCAAACGCCGCGCTGTTGATTCAGCCGTTTGA 553
255 aGlyCysValGluAsp.....ThreugluAlaIleAla 266
554 CCGAAGCCAAATTCATGTTTGTAGGACGTCGCGACAGATGCCGCTC 603
266 laGlyTyrThrIleHisThrPheHisThrGluGlyAla..... 278
604 GAAATGCTGCCAATCGAAGACATGATTCGCGCGCCGCGATCTGC 653
279 .....GlyGlyGlyHisAlaBr 284
654 CCGTTTGAGTGGCAGCAGCATTCATTC...ATCGAGCCGTCGCGCGCA 700
284 oAspValIleLysMetAlaGlyLysPheAsnIleLeuProAlaSerThrA 301
701 ATAAACCGTG...TGACCATCAT.....TATCAAGATGTA 735
301 snProThrIleProPheThrLysAsnThrGluAlaGluHisMetAspMet 317
736 ATTAACATTCGCGCT.....TTGTT 755
318 LeuMetValCysHisHisLeuAspLysSerIleLysGluAspValGlnPn 334
756 TGCAACAGGCGCTGGAACACCGCGCGATGCGCCCTAGTGCTTC 805
334 eAlaAspSerThrGlyLeuArgProGlnThrIleAlaAla.....GluAspG 349
806 AAGTCACAAACCGCGCTCTTGCGTACCGTTGGTGGCGAAAGTATCG 855
349 ILeuHisAspMetGlyLePheSerIleThrSerAspSerGlnAla 365
856 CAATTCAGCGCGCGATTCGTT..... 879
366 MetGlyArgValGlyLysValIleThrArgThrTrpGlnThrAlaAspLy 382
880 .....GACACAGACA 889
382 sAsnLysLysGluPheGlyArgLeuLysGluGluLysGlyAspAsnAspA 399
890 AC.....CGCGTATTCGCGTTCGGATTCGAACGGCGCAT 927
399 snPheArgLysArgTyrIleSerLysTyrThrIleAsnProGlyIle 415
928 ACACAGAGGCGCGCATATTCGAGCGTACACCATCAGATTTCGCT 977
416 AlaHisGlyIleSerAspTyrValLys..... 425
978 TATCGAAGAAGCGCGACGACCAAGACCTGCGCTGC..... 1014
426 .ValGluValGlyLysTyrAlaAspLeuValLeuThrSerProAlaPheP 442
1014 ..... 1014
442 heGlyLysProAsnMetIleLysGlyGlyPheIleAlaLeuSer 458
1015 .....GTTCGCCGCGACCGGACCAATA 1037
459 GluMetGlyAspAlaAsnAlaSerIleProThrProGlnProValTyrTY 475
1038 CTCATTCAGCGGTACAACCTCGCGCATTCCTGAAAAACAACCTCTTCA 1087

```

```

475 r.....ArgGluMetPheGlyHisGlyLysAsn.....L 486
1088 AGTTCACACAGCCCGTCAAC..... 1107
486 yAsPheAspThrAsnIleThrPheValSerGlnAlaIaTyrLysAlaGly 502
1108 .....GGCGGCGACCGCGCATGTCGCGATTCGCTACTTA 1142
503 IleLysGluGluLeuGlyLysLeuAspArgAlaAlaProProValLysAsnCy 519
1143 CGAGCGCGTATGCCCTTCGATATC 1167
519 sArgAsnIleThrLysAspLeu 527

seq_name: /cgn2/6/ptodata/1/aa/5B_COMB.pep:US-08-564-972-1
seq_documentation_block:
; Sequence 1, Application US/08564972
; Patent No. 5843462
; GENERAL INFORMATION:
; APPLICANT: Cont-Fine, B. M.
; TITLE OF INVENTION: DIPTHERIA TOXIM EPITOPES
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/564,972
; FILING DATE: 30-NOV-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Karen D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.344051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-564-972-1

alignment_scores:
Quality: 94.50 Length: 489
Ratio: 0.450 Gaps: 28
Percent Similarity: 42.945 Percent Identity: 20.245

alignment_block:
US-09-303-518d-125 x US-08-564-972-1 ..
Align seg 1/1 to: US-08-564-972-1 from: 1 to: 535

```

```

10 ATCAAAAAAGGTCTAAACCTGCCATCCGGCGACAGCCGACGCGT 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 Ileglnlysglylleelnlyspolyseryl.....thrglnlyas 45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 TTACGAC.....66
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
45 ntyraspaspaprtprlysglyphetyrserthraspanslytyraspa 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 .....GGCCGGCCATT...ACCAAGTCGGCTGCTGGCGAA..... 102
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 laalaglytyrservalaspasnlnuansprouserglylysala gly 78
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 .....GAATAGCCGATGCGGCC.....TCGATGAA 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 glyvalalysvalthrtyrproglyleuthrlysalenualenly 95
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 AGTCAGGAGAGCGATGCCGTCACAAAAAGCCAGTGTGTAAGACA 181
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 svalaspasnlaeluthrile.....L 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 AAAAGAAATCCGGCGGTGTACTAGCGCGGCTTCAGGCAAAATCGCC 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 ysllysglnleuglyleuserleuthrleuhenetcluglnvalgly 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 GCG.....ATCACCGT.....GGCAAAAGCGGTACTTCA 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 thrcluglnphelelysarpheneglyaspglyalaserargvalalle 136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 GTCAGTCGATGTCGGTGTGAGGCAAGC.....294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 userleupropheala...glulyserserservalglutryleasna 152
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....GAATAGGATTTGAA 309
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 sntrpelnlnalalysalaleuservalglulnleu lnleasnhegn 168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 .....CGCTACGACCTGAAGC 326
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 thraarglylsararglylnasplametyrglutrymetlalnlnal 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 GCTGGCAAACTTAAGCGGGAAGATGGCGCAACTGATTCACACCG 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 acysala.....glyasnargvalargyser...Valglysers 198
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
377 GTTG.....TGACTGCGCTCGCACCCGCTCCGCTC 408
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 erleusererylleasnleuansprrpasvallleargasplysThrlys 214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
409 AGCAAAATTCCTGCGGTGATGCCGCGGCTTCGCAATCTTCGTCATGC 458
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 thrlyslleuglyserleu.....lysluhslelyprolelyshanly 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 GATGGACACCAATCGCTGCGTGGCGGACCCATGATTAATCAAGAG 508
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 shetserleuserProasnlysrThrValserglululysalalyseln 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
509 CGCGCGAGATTTCAACGCGGCTGTGATTAAGCCGCTTGACCGAA 558
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 yrluengluluphehslethrhalaleu lnhsproglu leusergl 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
559 CGCAAAATCCATGTTTGAAGCAGCTGGCGCACAGCTGCCGCTGAAAA 608
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 leulysThr.....265
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 TGCTGCCAACATCGAAACACATGAATTCGGGCGGCCCATCTCGCGGT 658
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 .....V 266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 TGAATGGACGACATTCATTCATCGAGCGCGTGGCGGAATAAACC 708
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 altnhglythrAsnProvalphe.....AlaglylaAsenryala 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
709 GGTGACCAATCAATTATCAAGATGTAATACC.....741
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 AlatrpalaValasnvalAlaglnvalleaspserclunthralaspas 296
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 .....ATTGGCGCT 751
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 nleuglnlysrthrhalalaleuserlleleuproglylle glyser 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
752 TG.....TTTGCAACAGCCGCTCTG.....AACACGAGCGCGTGA 789
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 almetglyllealaasp glyalavala lnhsasnthrcluglnleval 329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 GCCCTAGGTGGTCTCAAGTCACAAACCGCGGCTCTGCGTACCGTT 839
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 Ala.....lnserllealaleuserleu 339
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 GGTGCAAAATATCGCAATTAATCTCGGCGCAATGTTGACACAGACA 889
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 tvalAlaglnlnalleproleu...ValglylnleuValasp..... 352
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
890 ACCGCGATTCGCGTGGTGTATGACGGCGCATTCACAGCGCG 939
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 .....llelyphealala 357
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
940 CAGCATTAATTTGGACGCTACCAATCAGATTCGTTATGAGAGAG 989
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 tyrasnphelvalgluserlleleasnleupheneglnval lnhsasnse 374
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 CCCGACCAAGACCTGTCGGCTGGGCTGGCGG.....CAGC 1027
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 tTyraAsnArgProAlaTyr.....SerProdynhslysrthrglnp 388
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1028 CG.....GACAAATCTCATCAGCGGTACCAACCTGCGCATTTTC 1068
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 roheleu lnhsasp glytyrAlaValaserThrpsanThrValgluAspser 404
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1069 CTGAAAAACAACCTCTCAAGTTCAACACAGCCGCAACGCGCGCACCG 1118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 lle.....lleargthrlypheneglnlyglusergl 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1119 CGGCATGTCGCGATTTGATGACTTACAGAGCGGTGACCTTGAT.... 1164
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
415 ylnhsapllelysrlethrAlaValasnthrProleupro lleAla gly 432
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1165 ..ATCTGCCACCTG 1179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
432 al leu leuproThrle 437
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/plodata/1/laa/6B_COMB.pep:us-09-171-969-9
seq_documentation_block:
: Sequence 9, Application US/09171969
: Patent No. 6284533
: GENERAL INFORMATION:
: APPLICANT: Thomas, Lawrence J.
: TITLE OF INVENTION: PLASMID-BASED VACCINE FOR TREATING ATHEROSCLEROSIS
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Witcoff, Ltd.
: STREET: 75 State Street, Suite 2300
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1807
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Wordperfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/171,969
: FILING DATE: 01 May 1997 (01.05.97)
: CLASSIFICATION: 514

```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/640,713
 FILING DATE: 01 May 1996 (01.05.96)
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/802,967
 FILING DATE: 21 February 1997 (21.02.97)
 ATTORNEY/AGENT INFORMATION:
 NAME: Leon R. Yankwich
 REGISTRATION NUMBER: 30,237
 REFERENCE/DOCKET NUMBER: TCS 414.1 PCT (05872)
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 535 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL:
 ANTI-SENSE:
 FEATURE: (A)NAME:
 LOCATION:
 US-09-171-969-9

alignment_scores:
 Quality: 94.50 Length: 489
 Ratio: 0.450 Gaps: 28
 Percent Similarity: 42.945 Percent Identity: 20.245

alignment_block:
 US-09-303-518D-125 x US-09-171-969-9 ..

Align seg 1/1 to: US-09-171-969-9 from: 1 to: 535

```

10 ATCAAAAAGCTTAACCTGCCATCGCGGAGACCGAGACCGCT 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
31 Ileglnysgylleglnyspirolyseryl.....Thrglnlyas 45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
60 TTACGAC..... 66
45 ntyraspsasprtrprrysglyrhetysertlnaspsnlystyraspa 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
67 .....GGCCGCGCATTT...ACCGAGTCGCGTTCCTGCGCAA..... 102
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
62 laalaglytyrservalaspsnnglnasnprouleuserglylsalagly 78
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
103 .....GAATATGCCGATGCGCC.....TCGATGAA 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
79 glyvalvalylsvalthrtprrogllyneuthrlysalenaleu 95
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
132 AGTCAGGAGAGCGATCGCTCAAAAAAGCCAAAGCTCTTTGAGACA 181
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
95 svalaspsnlaaglnthrile.....L 103
187 AAAAATCCGCGCGCTGCTTACTGCGCGGCTTCAGGAAATAGCC 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
103 yslgsluenglyleuserleuthrlnprouleuserglylnvalagly 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
120 thrnglnuphellelyargpneglyaspilyalaserargvalvalle 136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
264 GTCAGTCGATTCGCTGTTGAAGCAAGAC..... 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
136 userleupropheala...Glnglysersestervaleglutyrilleasna 152
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....GAATGAGCTTTGAA 309
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
152 sntrpoglunlalalysalaleuservalgluengluilleasnphel 168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
310 .....CGCTACGCACTGAGC 326
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
169 Thrargglylsgargglylnaspsalameetyrglutymetalaaglnal 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
  
```

```

327 GCTGGCAAACTTAAGCGCGAAGATGCGCGGCAACCTGATTCATCG 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
185 aaysala.....Glyasnargvalargser...Valglyers 198
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
377 GTTTG.....TGAGTCGCGCTGCGGACCGCGCTTC 408
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
198 erleusercylleasnleusprtrpaspvalilleaspslysthrlys 214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
409 AGCAAAATCTCGCGCTGCGGACCGCGGCTTCGTAATGC 458
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
215 thrlgslleugluserleu.....Lysglunhsglyprolileysasly 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
459 GATGACACCAATCCGCTGCGCTGCGGACCTACGCTATTCAAAAGAG 508
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
229 smetsergluserproasnlsthrvalserglunlulysalalysglnt 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
509 CCGCCGAGATTTCAACGCGCGCTGTTGATGAGCGTTTGACGCA 558
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
246 yrluengluphehnsnglnthralaleuglnhlsprogluuserglu 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
559 CGCAAAATCCATGTTTGTAAAGCAGCTGCGGACGAGCTGCGCTGA 608
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
263 leulysthr..... 265
609 TGCTGCCACATGAAACACATGATTCGCGCGCGCATCTCGCGGT 658
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
266 .....V 266
659 TGAGTGGACGACATTCATTCATGAGCGCGGCGGCAATTAAC 708
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
266 althrglythrpsnprovalphe.....Alaglylalsnlyrha 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
709 GTGTGACATCAATTAATCAAGATTAATACC..... 741
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
280 Alatrplavalasnlvalalaglnvalilleaspserylthralasps 296
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
742 .....ATTGCCGCT 751
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
296 nleuglnlysthrthralaleuserlleuprogllylleglyserv 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
752 TG.....TTTGCACAGCGCGCTG.....AACCCGACCGCTGAT 789
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
313 almetglyllealaspglyalavalhnsnsthrglnuglnleval 329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
790 GCCCTAGCTGTTCTCAAGTCACAAACCGCGCGCTGCGTACCGCTTT 839
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
330 Ala.....Glnserllealeuserleu 339
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
840 GGTGCGAAGATATCGCAATTAATCTGCGGCGAATGTTGACACAGACA 889
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
339 tvalalaglnalaleuproleu...Valglyluleuvalasp..... 352
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
890 ACCCGGTGATTCGCGTTCGATGAAAGCGCGGCTTACACAGCGCG 939
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
353 .....lleglyrhealiala 357
940 CACGATTTTGGAGCTACACACATGATTCGCTTATTCAGACAGAG 989
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
358 tyrasnphelvalgluserllelleasnleupheglvalvalhnsnse 374
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
990 CCGCAGACAGAGCTGTTGCGCTGCGTTCGCGG.....CAGC 1027
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
374 rtyrasmnargproalaty.....Serproglunhlslysthrlnb 388
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1028 CG.....GACAAATCTCATCAGCGGCTCAACCCCTCGGCAATTC 1068
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
388 tophleunhlsaspllytravalalasertrpsntrhvalguaspser 404
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1069 CTGAAAAACAAACCTTCAGATTCACACAGCGCTCACGCGCGGACG 1118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
405 ile.....lleargthrcllypneglnglylusergl 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1119 CGCCATGCTGCGGATTCGCTTACGAGCGCGGTATCCCTTGAT.... 1164
  
```


seq_name: /cgn2_6/plodata/1/aa/5B_COMB.pep:US-08-760-489-2

seq_documentation_block:
Sequence 2, Application US/08760489
Patent No. 5830696

GENERAL INFORMATION:
APPLICANT: Short, Jay M.
TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-08-760-489-2

alignment_scores:
Quality: 94.50 Length: 547
Ratio: 0.475 Gaps: 27
Percent Similarity: 36.380 Percent Identity: 20.658

Alignment_block:
US-09-303-518D-125 x US-08-760-489-2 ..

Align seg 1/1 to: US-08-760-489-2 from: 1 to: 1487

```
15 AAAGGTAAACCTGCGGCGGAGCGGAGCGGAGCGGCGTTTACG 64
   :::::::::::::::::::: |||:::
295 ArgGlnArgProAlaArgGlyLysPlySglnLeuArgThrLeuAl 311
   :::::::::::::::::::: |||:::
65 ACGG.....CCGGCCATTACCGAAGTCGGTGGT.....T 96
   |||::: |||::: |||:::
311 gGlnGlySerProArgGlnArgProArgSerGlnArgLeuGlnAsp 328
   |||::: |||::: |||:::
97 GCGGAAGATATCGCGGTATGCGCCCGCGATGAAGTCAGGAAGCGA 146
   |||::: |||::: |||:::
328 rPasnArgVal..... 331
147 TGCCTCAAAAAAGCGCAGTCTGTTGAAGACAAAAAGATCCGGCG 196
   :::: |||::: |||:::
332 .....GlnGlnAspLeuSerLeuAl 338
```

```
197 TGCTGTTACTGCGCGCGC.....TTCA 219
   :::: |||::: |||:::
338 AsnValValLysGlyLysArgAlaGlyLysLeuArgThrArgGlyVal 355
   :::::::::::::::::::: |||:::
220 GCGAAATCGCGCGGATTCACCGTGGCGAAAGCGCGTACTTCAGTCAGT 269
   |||::: |||::: |||:::
355 rGlnAsnArgArgHisAla.....ArgArgAlaArg..... 365
   |||::: |||::: |||:::
270 CGTGAATCCGTTGAAGCGAAGCGAATTCGATTTGAACGCTACGAC 319
   |||::: |||::: |||:::
366 ArgAspSerGlu..... 369
320 CTGAAGCGCTGGCGAA.....CTTAAGCGCGAAGATCGCGCGCA. 362
   |||::: |||::: |||:::
370 ....SerSerGlyAspSerLeuLeuProArgTyrAlaArgProGln 385
   |||::: |||::: |||:::
363 .....CCTGATCCCAATCCGGTTGTGTGAC 386
   |||::: |||::: |||:::
385 LValGlyLeuGlnGlnHisArgGlnProGlnProLeuHisAlaProPro 401
   |||::: |||::: |||:::
387 TCGCGTGGCGCGACCG..... 401
402 LeuAlaSerArgSerAspAsnArgGlnGlyGlnGlyProHisGlnArgAs 418
   |||::: |||::: |||:::
402 .....TCCG 405
402 pTrpLeuGlyArgAlaGlyLysArgGlyLysValArgGlnValArgGly 435
   |||::: |||::: |||:::
406 TTCAGCAAAATTCCTGCGCGTGCATGCGCGCGTTCGATCTCGTCAA 455
   |||::: |||::: |||:::
435 aHisArgGlnArgThrArgGlyProArgTyrValGlnHisLeuGlnArg 451
   |||::: |||::: |||:::
456 TCGCATGACAC.....CA 469
   |||::: |||::: |||:::
452 AlaAspGlyArgGlyLysAlaArgLeuProArgAlaLeuLeuArgLeuSe 468
   |||::: |||::: |||:::
470 ATCCGCTGCGTGGCGACCGCTACGGT.....CAT 498
   |||::: |||::: |||:::
468 rAlaGlyGlyTyrGlnProArgGlyGlyLysAlaGlnAsnProGlnHisA 485
   |||::: |||::: |||:::
499 ATCAAGAGAGCGCG..... 512
   |||::: |||::: |||:::
485 sPlysArgProArgThrGlyLeuGlnAspAspLysGlnValArgGlnGly 501
   |||::: |||::: |||:::
513 .....CGAGATTTCAAAGCGCGCTGTGGATGACCGCTTGACCG 556
   |||::: |||::: |||:::
502 LysGlyArgGlyPheProLeuArgGlyArgGlyArgAspAsnLeuGlnG 518
   |||::: |||::: |||:::
557 AACGCAAAATCCATGTTTGAAGCAGCTGGCGCAGAGCTGCCGCTGAA 606
   |||::: |||::: |||:::
518 nHisArgArg...CysLeuSerLeuArgLeuGlnArgProLysGlyArgG 534
   |||::: |||::: |||:::
607 AATGCTGCGCAATCGAAGAACACATGAATTCGGCGCGCGCATCCGCG 656
   |||::: |||::: |||:::
534 LysGlyArgGlyArgGlnLeu...LeuProGlnArgAlaLeuLeuArg 549
   |||::: |||::: |||:::
657 TTTGAGTGGCAGCGACATTCATTCATTCAGCGAC..... 689
   |||::: |||::: |||:::
550 ArgAsnProGlnGlnAlaGlnHisArgValArgArgAspLeuArgGln 566
   |||::: |||::: |||:::
690 .....GCTGCGCGCGAATTAACCGCTGTGACCAATCAAT 723
   |||::: |||::: |||:::
566 nSerSerAlaSerGlnGlyLysArgLeuAspArgValLeuLeuHisGlnA 583
   |||::: |||::: |||:::
724 TATCAAGATGTAATTAC.....CATTCGCGCTTTGTT 755
   |||::: |||::: |||:::
583 rSerArgGlnValPheGlyLysAlaGlnValProGlnHisThrProAspIle 599
   |||::: |||::: |||:::
756 TGCACACAGCGCGTGTGAACACGAGCGCGTGATTCGCTAGTGCTTC 805
   |||::: |||::: |||:::
600 LeuProGlySerSerGlnLeuArgLeuArg..... 609
   |||::: |||::: |||:::
806 AAGTCAACAACCGCGCTTGGGTACCGTTTGGGTGCGAAGATATCG 855
```

```

609 ..... 609
856 CAATTACTGGGGCGAATTGTTGACACAGACACCGCGTATTCGG 905
610 ..... LeuGlnAlaArgGluPhe... 615
906 TTCGGTATTGAAGCGCGATTTACACAAAGCGCGACGATTTATTCGG 955
616 ..... PheArgArgArgGlyAlaArgGlySerHisArgLeu... GlyAs 629
956 GCTACCAATCAGATTCGTTATTCAGACAGCGCGACGCGAGAGCTG 1005
629 PLeuSerGlyGlyAspLeuArgLeuAspGlyArgGlyGln... I 644
1006 TTCGGTGGTGGCGCGACGCGGACAAATATCTCATCAGCGGTACAC 1055
644 LeArg.....Gly 646
1056 CCTGGCCATTTCCTGAAAAAACAACCTCTCAATTCAACAGCGCTCA 1105
647 Pro.....GlyLeuArgHisArgGlySer 654
1106 ACAGCGCGACCGCGCATGTGCGCATGTGACTTACAGCGCGGTATG 1155
654 GAsnSerArgPheAsnHis.....ProA 662
1156 CCCTTGGAATTCCTGCCACCCCTGTTTGGCGATTTATTCGTCGGGA 1205
662 LaAlaValLeuProArgGluProGlySerGlyAlaArgGly 678
1206 TACCGACAGCGCGACGATTTGGGTTGCTTGAATTTGACGAGA... 1250
679 GlyLeuArgArgGlnGlyLeuProLeuGlnGlyAlaAspArgGlnLeuArg 695
1251 ....AGACCTCGCTTGTGCAGCTTCGTCTGCCGCGGAAATACGAATAC 1296
695 GValGlyProArgPhe.....GlnAspGluValA-705
1297 GCGCCGCTGTCGCAAGCTGTGGAACATTGAGAGGA 1337
705 rGProLeu.....SerGlySerHisAsnGlnGly 714

seq_name: /cgn2_6/plodata/1/laa/58_COMB.pep:US-08-760-489-4
seq_documentation_block:
: Sequence 4, Application US/08760489
: Patent No. 5830696
: GENERAL INFORMATION:
: APPLICANT: Short, Jay M.
: TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: US
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: IBM Compatible
: SOFTWARE: FASTSEQ for Windows 95
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/760,489
: FILING DATE: 05-DEC-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/008,311
: FILING DATE: 07-DEC-1995
: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Haile, Ph.D., Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/008001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5090
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1487 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: US-08-760-489-4

```

```

alignment_scores:
  Quality: 94.50      Length: 547
  Ratio: 0.475        Gaps: 27
  Percent Similarity: 36.380  Percent Identity: 20.658

```

alignment_block:

US-09-303-518d-125 x US-08-760-489-4 ..

Align seg 1/1 to: US-08-760-489-4 from: 1 to: 1487

```

15 AAAAGTCTAAACCTGCCCCATCGCGGACAGCGGAGCAGCGCTTACG 64
295 ArgGlnArgArgProAlaArgGlyGlyAspGlyGlnLeuArgThrLeuArg 311
65 ACCG.....CCGGCATTTACCGAAGTCGCTGCT.....T 96
311 gGluGlySerProProArgGlnArgProArgSerGlyArgLeuGlnAsp 328
97 GCGAAGAATATGCGCGTATGCGCCCTCGATGAAAGTCAAGAGAGCGA 146
328 rPAsnArgVal..... 331
147 TCCCGTCAAAAAAGCGCAAGTGTGTTGAAGACAAAAAGATCCGGCG 196
332 .....GluGlnAspLeuSerLeuAl 338
197 TGGTGTATTACGCGCGCGC.....TTCA 219
338 AsnValValCysGlyGlyArgAlaGlyGlnLeuArgThrArgGlnGlyA 355
220 GCGAAATCGCCCGCATTCACCGTGGCGAAAGCGCTACTGAGTCAGT 269
355 rGlnAsnArgArgHisAla.....ArgArgAlaArg..... 365
270 CGTGATTCGCGTTGAAGCAACGCAAAATCGAGTTGAACGCTACGCAC 319
366 ArgAspSerGlu..... 369
320 CTGAAGCGCTGGCAAA.....CTTAAGCGCGGAAGAAGTCCGCCCAA 362
370 .....SerSerGlyAspSerLeuLeuProProArgGlyArgAlaArgGln 385
363 .....CTGATCCAATCCGGTTTGTGAC 386
385 lYalaglyLeuGlnGlyHisArgGluProGlnProLeuHisAlaProPro 401
387 TGCCTCGCGACCGG..... 401
402 LeuAlaSerArgSerAspAsnArgGluGlyGlyProHisGlnArgAs 418
402 .....TCG 405
418 pThrLeuGlyArgAlaGlyGluArgGlyGlyValArgGlnValArgGly 435
406 TTCAGCAAAATTCGCGCGCATGCGCGAGCGCTTGCCATTTGCTCAA 455
435 AlHisArgGluArgThrArgGlyProArgGlyValGlnHisLeuGlnArg 451

```

```

456 TGGATGACAC.....CA 469
457 AAlAspGlyArgGlyValAlaArgLeuProAlaAlaLeuLeuArgLeuSe 468
470 ATCCGCTGGCTGGCCAGCCCTACGGT.....CAT 498
468 rAlaGlyGlyTyrGluProArgGlyGlyValAlaGlyAsnProGlnHisA 485
499 ATCAAGAACCCG..... 512
485 spLysArgProArgTyrGlnGlyLeuGlnAspAspLysGlyValArgGlnGly 501
513 .....CGAGATTTCAACGCGCGCTTTGGTATTGAGCCCTTGACCG 556
502 LysGlyArgGlyPheProLeuArgGlyValArgGlyArgAspAsnLeuGlnG 518
557 AACGCAAAATCCATGTTTGAAGGAGCTGGCGGAGAGCTGCCGTCTGA 606
518 nHisArgArg...CysLeuSerTLeaArgLeuGlnArgProLysGlyArgG 534
607 AATGCTGCAACATCGAAACACATGATTGCGCGCCGCTGCTGCGCG 656
534 LuserGysArgLysArgGlnLeu...LeuProGlnArgAlaLeuLeuArg 549
657 TTGAGTGACACGACATTCATTTCATCGAGC..... 689
550 ArgAsnProGlnGlyGlnAlaGlnHisArgValArgArgAspLeuArgG 566
690 .....GGTCGCGCGCAATGAACCGTGTGACACATCAAT 723
566 nSerSerAlaSerGlnGlyGlnArgLeuAspArgArgLeuGlnGlnG 583
724 TATCAAGATGTAATAC.....CATTGCCGCTTGT 755
583 rGserArgGlnValPheGlyAlaGlnValProGlnHisThrProAspIle 599
756 TGCAACAGCGCTGTGAACAGCGCGTGAATGCGCTAGTGTTCTC 805
600 LeuProGlySerSerGlnLeuArgLeuArg..... 609
806 AAGTCACAAACCGCGCCTTGGCTACCGTTGGTGGCGAAAGTATCG 855
609 ..... 609
856 CAATTAATCGCGCGGATTTGGTGAACAACAACCGCGATTTCCG 905
610 .....LeuGlnAlaArgGlnPhe... 615
906 TTCGATATGAAGCGCGGATTTACAAGCGCGCAGATTTATTTGGAC 955
616 .....PheArgArgArgLysAlaArgLysArgHisArgLeu...GlyAs 629
956 GCTACCAACATCAGATTCGTTATCGAAGAAGCGCGCAGCAAGAGCTG 1005
629 pleuSerGlyGlyAspLeuArgLeuAspLysArgGlyGlnGln.....I 644
1006 TTGCGCTGGGTGGCGCGAGCGGCAAAATCTCCATCAGCGCTACAC 1055
644 LeArg.....Gly 646
1056 CCTCGGCGCATTTCTGAAAACAACCTTCAAGTTCAACACAGCGCGCA 1105
647 Pro.....GlyLeuArgHisArgLysArg 654
1106 ACGCGCGGACCGCGCATGGTGGCGATTTGACTTACAGCGCGCTGATG 1155
654 gAsnSerArgPheAsnHis.....ProA 662
1156 CCGTTGATATCTCGCCACCGCTTTTGGCGATTTAATCTGCGCGCA 1205
662 LalaValLeuProArgGlnProCysSerGluAspGlyGlyValArgGly 678

```

```

1206 TACGACAGCGCGAGCATTTGGTGTCTTGATTTGACAGAGA..... 1250
679 GlyLeuArgArgGlnGlyLeuProLeuLeuGlyAlaAspArgGlnLeuArg 695
1251 .....AGACCTGCTTTGTGACAGCTTCTGTCGCGCGCAATATGAAATAC 1296
695 gValGlyProArgPhe.....GlnAspGlyValA 705
1297 GCGCGCTGTTGCGCAAGTGTGGAACATGGAAGA 1337
705 rGProLeu.....SerGlySerHisAsnGlnGly 714

```

```

seq_name: /cgn2_6/ptodata/1/aa/6B_COMB.pep:US-09-185-373-2

```

```

seq_documentation_block:
/ Sequence 2, Application US/09185373
/ Patent No. 6335179

```

```

GENERAL INFORMATION:
APPLICANT: Short, Jay M.
TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
ENZYMES

```

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/185,373
FILING DATE: 03-NO. 6335179-1998
CLASSIFICATION: <Unknown>

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Halle, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5099
TELEFAX: 619/678-5070

```

```

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

```

US-09-185-373-2

```

```

alignment_scores:
Quality: 94.50 Length: 547
Ratio: 0.475 Gaps: 27
Percent Similarity: 36.380 Percent Identity: 20.658

```

```

alignment_block:
US-09-303-518D-125 x US-09-185-373-2 ..

```

```

Align seg 1/1 to: US-09-185-373-2 from: 1 to: 1487

```

```

15 AAAGGCTTAACCTGCCATCGCGGAGACCGGAGCAACCGCTTACG 64
.....

```


295 ArgGlnArgProAlaArgGlyLysPlySglnLeuArgThrLeu 311
65 ACG.....CCGGCATACCGAAGTCGGTCT.....T 96
311 glnGlySerProAlaArgGlnArgProArgSerGlnArgLeuInAsp 328
97 GCGGAGAAATGCGGATATGCGCCCTCGATGAAGTCAAGGAAGCGA 146
328 rPasnArgVal..... 331
147 TCCCGTCAAAAGGCCAAGTCGTGTTGAAGCAAAAGATCCGGCG.196
332GlnInAspLeuSerLeuAl 338
197 TGGTGTATTACTGCGCCGCG.....TTCA 219
338 aasnValValCysGlyArgAlaGlnLeuArgThrArgGlnGlyA 355
220 GCGAAATCGCGCGATTCAACGTCGGGAAAGCGGCTACTTCAGTCAGT 269
355 rGlnAsnArgArgHisAla.....ArgArgAlaArg..... 365
270 CGTATTCGCGTTGAAGCAAGCAAGATCGATTGAACGCTACGAC 319
366 ArgAspSerGlu..... 369
320 CTGAAGCGCTGGCAMA.....CTTAAGCGCGAAGAGTCGCGCGCA. 362
370SerSerGlyAspSerLeuProProAlaGlyArgAlaProGlnG 385
363CTGATCCAATCGGTTTGTGAC 386
385 lylalGlyLeuGlnGlyHisArgGlyProGlnProLeuHisAlaProPro 401
387 TCGGCTGCGCAGCCG..... 401
402 LeuAlaSerArgSerAspAsnArgGlnGlyProHisGlnArgAs 418
402TCG 405
418 pThrLeuGlyArgAlaGlyLysArgGlyGlyValArgInValArgGly 435
406 TTCAGCAAAATTCCTGCGTCGATGCGGCGGCTGCGCATTTGTCGCA 455
435 alHisArgGlnArgThrArgGlyProArgGlyValGlnHisLeuInArg 451
456 TGGGATGACAC.....CA 469
452 AlaAspGlyArgCysGlyAlaArgLeuProArgAlaLeuArgLeuSe 468
470 ATCCGCTGCGTCGCGACCTACGGT.....CATT 498
468 rlaGlyGlyArgGlnProArgGlyGlyLysAlaGlyAsnProGlnHisA 485
499 ATCAAGAAGCCG..... 512
485 sPlyAspArgProArgThrGlyLeuGlnAspAspGlyValArgGlnGly 501
513CGAGATTCAACGCGCGCTGTTGTTAAGCGGTTTGAACG 556
502 LysGlyArgGlyPheProLeuArgGlyArgGlyArgAspAsnLeuGlnG 518
557 AACGCAAAATCCATTTGTAAGCAGATGCGCAGACGCGCGCTCGAA 606
518 nhisArgArg.....CysLeuSerIleArgLeuGlnArgProLysGlyArgG 534
607 AATGTCGCAACATCGAAACACATGAATTCGCGCGCGCGATCCTGCGG 656
534 luserCysArgLysArgGlnLeu.....LeuProGlnArgAlaLeuArg 549
657 TTGAGTGGCAGCAGCATTCATTTCATCGAGCC..... 669
550 ArgAsnProGlnGlyGlnAlaGlnHisArgValArgArgAspLeuArgG 566

690GGTCGGCGGAAATAAACCGTGTGACCATCAT 723
566 nSerSerAlaSerGlnGlyLysArgLeuAspArgArgGlyLeuHisGlnA 583
724 TATCAAGATGAATTAC.....CATTGCGCGTTGTT 755
583 rSerArgGlnValPheGlyAlaGlnValProGlnHisThrProAspIle 599
756 TGCAACAGCGCGTGTGAACACCGAGCGCGTGAATTCCTAGTGTCTC 805
600 LeuProGlySerSerGlnLeuArgLeuArg..... 609
806 AAGTCAACAAACCGCGCTCTTGGTACGCTTTTGGGTGCAAGTATG 855
609 609
856 CAATTCATCGCGCGCAATTGGTTGACACAGACACACCGCGTATTTCCG 905
610LeuGlnAlaArgLysPhe... 615
906 TTCGATTGACGCGCGCATTCACAGAGCGCGCACGATTATTGGCAC 955
616PheArgArgArgLysAlaArgLysArgHisArgLeu...GlyAs 629
956 GCTACCAACATCAGATTTCGTTATCGAAGAGCGCGCAAGAGCGTG 1005
629 rLeuSerGlyGlyAspLeuArgLeuAspLysArgGlyGlnGln.....I 644
1006 TTGCGGCGGTTGCGCGCGACGCGCAAAATCTCCATCAGCGTACAC 1055
644 LeuArg.....Gly 646
1056 CCTGGCGCATTTCTGAATAACACTTTCAAGTTCAACACAGCGGCA 1105
647 Pro.....GlyLeuArgHisArgLysArg 654
1106 ACGCGCGCGACGCGCATGTCGCGATTGTTACGAGCGCGTGATG 1155
654 gaAsnSerArgPheAsnHis.....ProA 662
1156 CCTTGATATCTGCGCGCACCTGTTTGGCGGATTAAATCGCGCGA 1205
662 lalValAlaLeuProArgGlnProCysSerGlyAspGlyValArgGly 678
1206 TACCGACAGCGCGCAGCATTCGTTGTTGGAATTGGACGAGAGA... 1250
679 GlyLeuArgArgGlnGlyLeuProLeuLeuGlyAlaAspArgGlnLeuArg 695
1251AGACCTGCTTGTGACGCTTCGTCGTCGCGCGCAATACGAATAC 1296
695 gValGlyProArgPhe.....GlnAspGlyValA 705
1297 GCGCGCGTTGTCGCAAGTGTGAAACCATGAGAGGA 1337
705 rGProLeu.....SerGlySerHisAsnGlnGly 714

seq_name: /cgnr_6/prodata/1/iaa/6B.COMB.pep:US-09-185-373-4
seq_documentation_block:
: Sequence 4, Application US/09185373
: Patent No. 6335179
: GENERAL INFORMATION:
: APPLICANT: Short, Jay M
: TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
: ENZYMES
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: US


```
1006 TTGGCGTGGGTTGGCCCGGACGCGGACAAATCTCATTCACGGGTACAC 1055
      |||
644 leArg.....Gly 646
1056 CCTGGCCATTCTTGAAAAAACAATCTTCAGTTCACAGACGCGTCA 1105
      |||
647 Pro.....GlyLeuArgHisArgLysArg 654
1106 AGCGGGGACGCGCGCCATGTCGCCGATTGTACTTACGAGCGCGTGAAT 1155
      |||
654 GAsnSerArgPheAsnHis.....Pro 662
1156 CCCTGGATATCTGCCCCACCGCTCTTTGGCGGATTATATCGTCGCCCA 1205
      |||
662 laAlaValLeuProArgGluProCysSerGluAspGlyValArgGly 678
1206 TACCGACAGCGCGGACGCGATTGGGTGCTTGAATTCGACGAGA..... 1250
      |||
679 GlyLeuArgArgGlnGlnGlyLeuProLeuGlnGlyAlaAspArgGlnLeuArg 695
1251 .....AGACCTGCGCTTTGTGCGAGTTGTCGCGCGGCAATACGAAATAC 1296
      |||
695 GValGlyProArgPhe.....GlnAspGluValAla 705
1297 GCGCCGCTGTTGGCGCAAAAGTCTGGAACCATTCAGAAAGA 1337
      |||
705 rgrProLeu.....SerGlySerHisAsnGlnGly 714
seq_name: /cgn2_6/ptodata/1/laa/6B_COMB.dep:US-09-413-814-92
seq_documentation_block:
; Sequence 92, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, CO.
; APPLICANT: Bayer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 92
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-92
alignment_scores:
      Quality: 94.00      Length: 536
      Ratio: 0.423      Gaps: 30
      Percent Similarity: 41.418      Percent Identity: 22.015
alignment_block:
US-09-303-518d-125 x US-09-413-814-92
Align seq 1/1 to: US-09-413-814-92 from: 1 to: 608
28 CTGCCATCGGGGACAGACCGGAGCAAGCGTTTACGACGCGCGCCAT 77
1006 TTGGCGTGGGTTGGCCCGGACGCGGACAAATCTCATTCACGGGTACAC 1055
      |||
644 leArg.....Gly 646
1056 CCTGGCCATTCTTGAAAAAACAATCTTCAGTTCACAGACGCGTCA 1105
      |||
647 Pro.....GlyLeuArgHisArgLysArg 654
1106 AGCGGGGACGCGCGCCATGTCGCCGATTGTACTTACGAGCGCGTGAAT 1155
      |||
654 GAsnSerArgPheAsnHis.....Pro 662
1156 CCCTGGATATCTGCCCCACCGCTCTTTGGCGGATTATATCGTCGCCCA 1205
      |||
662 laAlaValLeuProArgGluProCysSerGluAspGlyValArgGly 678
1206 TACCGACAGCGCGGACGCGATTGGGTGCTTGAATTCGACGAGA..... 1250
      |||
679 GlyLeuArgArgGlnGlnGlyLeuProLeuGlnGlyAlaAspArgGlnLeuArg 695
1251 .....AGACCTGCGCTTTGTGCGAGTTGTCGCGCGGCAATACGAAATAC 1296
      |||
695 GValGlyProArgPhe.....GlnAspGluValAla 705
1297 GCGCCGCTGTTGGCGCAAAAGTCTGGAACCATTCAGAAAGA 1337
      |||
705 rgrProLeu.....SerGlySerHisAsnGlnGly 714
136 leuGlnLeuSerHisProAspLeuAlaIleThrValAspProHisAl 152
      |||
78 TACCGAATCGCGCTTGGCGCAAGAATATGCCGGATCGCGCCGTCA 127
      |||
152 alaMetLeuAla.....SerGluLysAspGlyLeuAlaIleThrValSer 168
128 TGAAGTTCAGAGAGGCGGATCGCTCAAAAAGCCCAAGTGTGTTGAA 177
      |||
168 yrlYsLeu..... 170
178 GACAAAAAGATCGCGCGTGTGTATTACGCGCGCTTCAGGCAAAAT 227
      |||
171 AspLeuHisGlyProSerValValGlnThrAlaCysSerHisSer 187
      |||
228 CGCGCGCATTCACCGTGGCGAAAGCGCGTACTTCAGTCAAGTCTG...A 274
      |||
187 uValAlaValHisMetAlaCysGlnSerLeuLeuGlnGlnCysAsp 204
275 TTGCGCTTGAAGCAAGCAAGCAATGAGTTGAAAGC..... 312
      |||
204 euAlaLeuAlaGlyGlyValSerIleGlyIleProGlnLysArgGly 220
313 ...TACGCACTGGAAGCGCTGGCAACTTACGCGCGCA..... 348
      |||
221 ProLysValProGlySerIleCysSerProAspArgGlyArgCysArgPro 237
349 .GAAGTGGCGCGGCAACCTGATCCAAATCCCGT..... 378
237 eaSprAlaArgAlaGlnGlyThrValGlyGlySerGlyValGlyIleVal 254
379 .....TTGTGACTGCGCGTGGCGCGCGCGCTG..... 408
254 laLeuLysArgLeuAlaSprAlaLeuArgAspArgAsnThrValHisAla 270
409 .....ACCAAAAT..... 417
272 ValIleArgGlySerAlaValAsnAspArgGlyArgGlyValGly 287
418 .....CCTGCGTGCATGCGGACGCGCTGCGCATC..... 447
287 emetAlaProSerValAspGlyGlnAlaAlaIleSerGlnAlaIle 304
448 .....TTGTCGCAATGCGCATGCAAC..... 465
304 euValAlaGlyValAspProGlySerIleGlyIleGlyValGlnAlaIle 320
466 ACCAATCGCGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515
      |||
321 ThrAlaThrAlaIleGlyAspPro...IleGlnValGlnAlaLeuThr 336
516 GGATTTCAAGCGCGCGCGTGTGATTCAGCGCGTTGACCGCAAGCAAA 565
      |||
336 naIaPheArg.....ArgLysThrProArgLys 346
566 TGCATGTTTGAAGCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCG 615
      |||
346 laArg...Cys...AlaLeuGlySer.....IleLysAla 355
616 AACATCGAAACATGCAATTCGCGCGCGCGCGCGCGCGCGCGCGCG 665
      |||
356 AsnIle...GlyHisLeuAspAlaAlaIleGlyValAlaGlnLeuIleLys 372
666 CACGCAATTCATTCATTCAGCGCGCGCGCGCGCGCGCGCGCGCG 712
      |||
372 laAlaHisValVal.....ArgSerGlyGluIleProCysVal 385
713 .....GACCCATCA..... 721
386 HisPheGlnAlaProAsnProLysLeuAspLeuAlaIleSerProPhe 402
722 .ATTATCAAGATGTAAATTCATTCGCGCG..... 749
      |||
```

```

402 eValProArGluAlaAlaPro.TPrProArGluLeuArGProArGArg 418
750 .....TTTGTTCACACAGCCGCTGTGACACCGAGCGG 784
419 AlaGlyValSerSerpheGlyIleGlyIleThrAsnAlaHisValIle 435
785 TGATTCCTCCTAGGTGGTCTCAAGTCACAAACCGCGCT..... 824
435 uGluGluProProLeuProLeuProArGAlaProAlaProGluArGAspH 452
825 .....CTTGGCTACCGCTTTGGTGGCGAAAGTATGCAATATAC 863.
452 lValLeuThrLeuSerAlaArGThrPro...GluAlaLeuSerThrAla 467
864 TGGGGCGAATTGGTTCACACAGACACCGCGCTGAT..... 899
468 CysAlaGluLeuAlaHisLeuGluAlaThrAspValProLeuAspAs 484
900 .....TTCCGGTTCGGTATTGACAGCGCGCATATACACAGCGCGCAC 942
484 pValAlaPheThrLeuGluThrGlyArGAlaHisProGlyArGArGAs 501
943 GATTATTGGAGCGCTACCAATCAGATTTCGGTATGCAAGAGCGCG 992
501 lAlaValAla.....ArgThrArGAlaGluAla 510
993 CAGCAAGAGCTGTGCGTGGTGGTGGCGCGACCGCAAAATATCTCA 1042
511 lIleGluLeuAlaArGAlaArGLeuGlyAlaSerAla.....LeuAl 523
1043 TCACGCGCTACACCCCTGGCCATTCTCTGAAAAACAACCTCTTCAAGTTC 1092
523 aArGProAspGluProArGPro..... 530
1093 AACACAGCGCTCAGCGCGCGACCGCGCATGTGCGCATGTGACTTCA 1142
531 ..SerSerArGSerAlaArGAlaArGArGProSerGlyTTP..... 544
1143 CGAGCGCGCTGATGCGCTGTGATATCTGCGCACCCCTGCTTTCCGGCAT 1192
545 .....Pro.....AlaArGSe 548
1193 TAATCGTCGGGATACGACAGCGCGAGCATGTGGTCTGTGCAATTG 1242
548 rThrArGArGArg.....ArgSerGlyAlaProSerThrArGAla 562
1243 GACGAGAAGACCTCGCTT.....TGTGCACTGTGCTGCC 1279
562 rGArGArGArgGlyArGAlaAlaSerIleSerAlaArGSerSerSerAla 578
1280 CGGCAATATACGATACGCGCGCTGTGGCAAAAGTGTGGAACCATT 1329
579 ArGAlaAlaGAlaThrGlyAlaArGcysSerAlaProArGTrpArGSerPr 595
1330 GAGA 1333
595 oArG 596

```

seq_name: /cgn2_6/ptodata/1/iaa/5b_COMB.pep.us-08-878-957-32

seq_documentation_block:

; Sequence 32, Application US/08878957

; Patent No. 5965796

; GENERAL INFORMATION:

; APPLICANT: Meagher, Richard B.

; APPLICANT: Summers, Anne O.

; APPLICANT: Kuhn, Clayton L.

; TITLE OF INVENTION: Metal Resistance Sequences and

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

```

CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,957
FILING DATE: 19-JUN-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/427,097
FILING DATE: 21-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Feider, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 40-94A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
TELEX:
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-957-32

alignment_scores:
Quality: 93.00 Length: 421
Ratio: 0.484 Gaps: 17
Percent Similarity: 45.606 Percent Identity: 20.190

alignment_block:
US-09-303-518D-125/rev x US-08-878-957-32 ..
Align seg 1/1 to: US-08-878-957-32 from: 1 to: 564

1122 GCGCGCGTCCGCCGCTGTGACGCGCTGTGGAACCTGGAAGAGTTGTTT 1073
1123 ..... 1124 .....
163 GlyIleAlaAlaAlaThrProThrIleGluAlaAlaAlaLeuAla 179
1072 TCAGAAATGAGC.....GAGGTTGTACCGGTGATGAG 1038
179 aGluGluGluAlaAlaArGValAspGluLeuArGHisAlaAlaTyrGly 196
1037 TATTTGTCGGCTCGCGCGCAACCCAGCGCATGAGTCTT.....GCT 994
196 le.....LeuGluGlyAsnProAlaIleThrValIleuHisGlySer 209
993 GCGGCTCTTCGATTAACGGAATCTGATGTGTA.....GC 956
210 AlaArGpHeuYsAspAsnArGAsnLeuIleValGluLeuAsnAspGly 226
955 GTCCCAATTAATCGTCCGCGCTGTGTAATCGCGC..... 919
243 roAlaValProProIleProGlyLeuYsAspThrProGlyTTPThrSer 259
918 .....GTTCAATACGCAACCGCA 901
900 AATCAACGGGTGTCTGT..... 883
260 ThrGluAlaLeuValSerGluThrIleProYsArGLeuAlaValIleG 276
883 ..... 883

```

```

seq_name /cgn2_6/ptodata/1/aa/5b_COMB.pep:US-08-804-227C-9
seq_documentation_block:
? Sequence 9, Application US/08804227C
? Patent No. 5876991
? GENERAL INFORMATION:
? APPLICANT: Dehoff, Bradley S.
? APPLICANT: Kuhstoss, Stuart A.
? APPLICANT: Kostock, Paul R., Jr.
? APPLICANT: Sutton, Kimberly L.
? TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
? NUMBER OF SEQUENCES: 15
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: THOMAS G. PLANT 1501
? STREET: LULU CORPORATE CENTER
? CITY: INDIANAPOLIS
? STATE: IN
? COUNTRY: USA
? ZIP: 46285
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: ASCII(DOS) Text only
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/804.227C
? FILING DATE: February 21, 1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Plant, Thomas, G.
? REGISTRATION NUMBER: 35,784
? REFERENCE/DOCKET NUMBER: X-8231
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 317-276-2459
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1996 amino acids
? TYPE: amino acid
? TOPOLOGY: unknown
? MOLECULE TYPE: peptide
US-08-804-227C-9

alignment_scores:
Quality: 93.00 Length: 471
Ratio: 0.472 Gaps: 23
Percent Similarity: 41.826 Percent Identity: 22.081

alignment_block:
US-09-303-518D-125 x US-08-804-227C-9 ..

Align seg 1/1 to: US-08-804-227C-9 from: 1 to: 1996

199 GTGTTTACTGGCGCGGCTTCAGGCCAAATGCSCGCAATTGACCGTGCGGA 248
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
749 VAISSerValAlaIalValaIasnglyProArgSerValValLeuSerGlyAs 765
249 AAAGCGGTACTTATGACGACGTCGCGATTGGCGTTGAAGGCAACGACGAAA 298
: ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
765 pValGIyValleuSuBerValValAlaSerLeuMetGlyASP..... 779
239 TCGAGTTTGAACGCTACGCACTGAAAGCGCTGCAAACTTAACGCGCGAA 348
80 .....gIyVal 781
349 GAATGCGCGCGCAACCTGATCSAATCCGGTTTTGCACTGCGCTGCGCAC 398
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
782 GLUCysArGtArGlyeudspValSerHisGlyPheHisSerValLeuMetGl 798
399 CGGTCCGCTTACGCAAAATTCCTCCGTC.....G 427
798 uPrOvalLLeugLySluPrheaGrGlyValValGluSerLeuGluPrheGlyA 815

```

```

428 ATGCCGAGCCGTTGCCATCTTGTCAATGCGATGAGCAACCAATCCGCTG 477
      |||||
815 rgalatrgproglvalvalvalvalseraserlserglglvalval 831
      |||||
478 GGTGCC.....GACCTTACGTCATTAATCAAGACCGCCGCA 515
      |||||
832 glyserglglleuglvalaspprogltytrtrvalarghisalaargl 848
      |||||
516 GGATTTCAAA.....CGCGGCTGTGGTAT 541
      |||||
848 valavalatrgprealaasrglvalglvalvalarglleuglval 864
      |||||
542 TGAGCCGTTTGACCGACGCAAAATCCAT..... 570
      |||||
865 ..gllythleuvalglvalglvalprohissglvalleuthglmetala 880
      |||||
571 ...GTTTGAAGCAGCTGGCGCAGAC.....GTCCGCTGTGAAA 608
      |||||
881 glglucylleuglvalglvalaspravalvalvalprolalametar 897
      |||||
609 TGCTGCGACATCGAAACACATGATTCGCGGC..... 642
      |||||
897 garglglvalglvalarglvalarglvalalalaleualathvalp 914
      |||||
643 ..CCGATCCTCCGCGTTTGAAGTGCACGACATTCATTCATCGAGCCG 690
      |||||
914 hethrargaspralaglglleualatathrthleunis...thglyser 929
      |||||
691 GTGCGGCGCAATTAACCGCTGAGCAGCATTAATCAAGATTAATAC 740
      |||||
930 thglvalarglglleaspleuothrtytrprohneghnis..... 943
      |||||
741 CATTTGCCGT.....TTGTTCAACAGCCGCTGTGACACCGCGCG 784
      |||||
944 ....Aspargtytrtrleualalaproserarproarthythasrglyl 959
      |||||
785 TGATGGCCCAAGTGTCTCAAGTCAACCAACCGCGCTGTGCGTACC 834
      |||||
959 euseralalaglglleualarglvalglvalproleuethrvalala 975
      |||||
835 GTTTTGGGTGCGAAAGTATCGCAATTAAGTGGCGGCAATTCGTTGACAC 884
      |||||
976 Val.....Gluleuprogllyth 981
      |||||
885 AGAACAACCGGTATTCGGTTCGGATTAAGACGCGCATTAACCAAG 934
      |||||
981 rasprthglvalatrrthlglarglleserlalaalasproulprolpr 998
      |||||
935 GCGCG...CAGGATTAATTTGGACGCTACCAATC..... 967
      |||||
998 eualalaspnis..leuvaltrpararglvalvalprogllythvalale 1014
      |||||
968 .....AGATTTCGTTATCGAAGAG 989
      |||||
1014 uleugluthvalleuglvalgllyserarglleugllyleuproarvala 1031
      |||||
990 CC..... 991
      |||||
1031 lagluleuvalleugluthrproleuthrtrthserasparproleu 1047
      |||||
992 .....GCGCAAAAGCTGTGGCTGGGTGGC 1020
      |||||
1048 glvalarglvalvalvalthralalalathralproglglvalaatr 1064
      |||||
1064 gluleuthrleunis.....serargprogluvalala..... 1076
      |||||
1071 GAAAAACAACCTTCAGAGTCAACACGCGGTACAGCGCGCGACCGCG 1120
      |||||
1077 .....AlaserSerSerProserProalaser.Proar 1088
      |||||

```

```

1121 CCAT.....GTGCGATTTGTAATTAACGCGCGCTG 1152
      |||||
1088 ghlsleuthralaglgluserasprasprrprrhrrghisalaserg 1105
      |||||
1153 ATGCCCTTGATATCCCGC..... 1172
      |||||
1105 lytleuvalalproalalaglglleualalaspasprhealagluleuthr 1121
      |||||
1173 .....CA 1174
      |||||
1122 glvalatrrprovalglvalaglulproleuaspleualagllylnty 1138
      |||||
1175 CCCGCTTTTCCGCGATTAATCGTCGCGATACCGACGCGCGCGCA 1224
      |||||
1138 rproleuphehalalaglglvalarg.....Tyrglgllyalalphea 1153
      |||||
1225 TTGGTTCCTTGGAATTTGACGAGAGAGACCTCGC...TTTGTGACGCT 1271
      |||||
1153 rglglleualarglalaatrrpararglglvaspglvalalphealaspval 1169
      |||||
1272 CGTGTGCGC 1280
      |||||
1170 Argleupro 1172
      |||||
seq_name: /cgn2_6/prodata/1/iaa/5B_COMB.pep:us-08-804-198-3
seq_documentation_block:
; Sequence 3, Application us/08804198
; Patent No. 5945320
;
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraia R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1996 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-198-3
alignment_scores:
Quality: 93.00 Length: 471
Ratio: 0.472 Gaps: 23
Percent Similarity: 41.826 Percent Identity: 22.081

```

alignment_block:

US-09-303-518D-125 x US-08-804-198-3

Align seq 1/1 to: US-08-804-198-3 from: 1 to: 1996

```

199 GTGTTTACGTGGCGGCTTCAGCAAAATCCGCGCATTCACCGTGGCA 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
749 ValSerValAlaValAlaValAlaValAlaValAlaValAlaValAlaValSerValValLeuSerGlyAsp 765
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 AAGCGCGTACTTCAGTCAGTGGTATTCGCGTGAAGCAACGACGAAA 298
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 pValGlyValLeuGlySerValAlaValAlaSerLeuMetGlyAsp 779
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 TCAGTTTGAACGCTACGACCTGAAGCGCTGCAAACTTAAGCGCGCA 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
780 .....GlyVal 781
349 GAAGTGGCGCGCACTGATCCATCCGCTTTGTGACTGCGCTGCGCAC 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
782 GluCysArgArgLeuAspValSerHisGlyPheHisSerValLeuMetGly 798
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 CCGTCCGTTACGAAATTCCTGCCGTC.....G 427
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
798 uProValLeuGlyGluPheArgGlyValValGluSerLeuGluPheGly 815
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 ATGCCGAGCGCTTCGCGCATCTTCATATGCGATGACACCAATCCGCTG 477
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
815 rGlyAlaArgProGlyValValValAlaSerSerValSerGlyValVal 831
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 GGTGCC.....GACCTACGCTCATATCAAGAAGCGCGCA 515
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 GlySerGlyGluLeuGlyAspProGlyTyrTrpValArgHisAlaArg 848
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 GAATTTCAAA.....CGCGCGCTGTGGAT 541
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
848 uAlaValAlaPheAlaAspGlyValGlyValAlaArgGlyLeuGlyVal 864
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 TGAGCGCTTGACCGCAACGCAATCCAT..... 570
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
865 ..GlyThrLeuValGluValGlyProHisGlyValLeuThrGlyMetAla 880
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
571 ...GTTTGAAGCGAGTGGCGCGAGC.....GTGCGCTGTGAAA 608
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
881 GluGlyCysLeuGlyAlaGlyAspValValValAlaProAlaMetAr 897
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
609 TGCTGCCACATCGAAACACATGAATTCGGCGGC..... 642
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
897 gArgGlyArgAlaGluArgGluValPheGluAlaAlaLeuAlaThrValP 914
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 ..CCGCATCTTGGCGGTTTGAGTGGCGACGACATTCATTCATGAGCG 690
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
914 heThrArgAspAlaGlyLeuAspAlaThrThrLeuHis...ThrGlySer 929
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
691 GTGCGCGGGAATAAACCGTGTGACCATCAATATCAATCATATTAATAC 740
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
930 ThrGlyArgArgLeuAspLeuProThrThrProPheGluHis..... 943
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
741 CATGGCGGT.....TTGTTGCAACAGCGCTGTAACCGCGAGCGG 784
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
944 ...AspArgTyrTrpLeuAlaAlaProSerArgProArgThrAspGlyL 959
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
785 TGATTGCCCTAGGTGGTTCAGTCAAGCAACAAACCGCGCTTGGCTAC 834
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
959 euSerAlaAlaGlyLeuArgGlyValGluHisProLeuLeuThrAlaAla 975
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
835 GTTTGGGTGCGAAAGATATCGCAAAATTAATTCGGCGGATTTGACAC 884
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
976 Val.....GluLeuProGlyThr 981
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
885 AGACAACCGCGTATTCGCGTTCGATTAAGCGCGGATTAACAAG 934
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
981 rAspThrGluValTrpThrGlyArgGlySerAlaAlaAspLeuProTrpL 998
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

935 GCGCG...CACGATTAATTGGACGCTACCAACATC..... 967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
998 euAlaAspHis..LeuValTrpAspArgGlyValValProGlyThrAlaLe 1014
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
968 .....AGATTTCGCTATGACGAGAG 989
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1014 uLeuGluThrValLeuGluValGlySerArgGlyLeuProArgValAla 1031
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
990 CC..... 991
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1031 IaGluLeuValLeuGluThrProLeuThrTrpThrSerAspArgProLeu 1047
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
992 .....GCAGCAAGAGCTGTGCGTGGCTGGC 1020
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1048 GluValArgIleValValThrAlaAlaAlaThrAlaProGlyGlyAlaAr 1064
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1021 CCGCAGCGCGCAAAATTCATCCATGACGCGTACACCGCTGCCATTTGCT 1070
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1064 gGluLeuThrLeuHis.....SerArgProGluProValAla..... 1076
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1071 GAAAAAACAATCTTCAAGTTCAACACAGCGCGTACAGCGCGGCGCGG 1120
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1077 .....AlaSerSerSerProSerProAlaSer.ProAr 1088
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1121 CCAT.....GTTGCCGATTTGCTATTACAGCGCGGTG 1152
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1088 gHisLeuThrAlaGluGluSerAspAspAspTrpThrArgHisAlaSerG 1105
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1153 ATGCCCTTGATATCTGCGC..... 1172
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1105 lYLeuLeuAlaProAlaAlaGlyLeuAlaAspAspPheAlaGluLeuThr 1121
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1173 .....CA 1174
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122 GluAlaTrpProValAlaGluProLeuAspLeuAlaGlyGlyTrp 1138
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1175 CCGTCTTGGCGGATTAATCTGCGCATTCGCGCGACGCGCGAGCA 1224
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1138 rProLeuPheAlaAlaGlyValArg.....TyrGluGlyAlaPheA 1153
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1225 TTGGTTGCTTGAATTGACGAGAGACCTGCG...TTTGGCGAGCTT 1271
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1153 rGlyLeuArgAlaAlaAlaTrpArgArgGlyAspGluValPheAlaAspVal 1169
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1272 GGTCTGCCC 1280
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1170 ArgLeuPro 1172
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /cgn2_6/prodata/1/1aa/5B_COMB.pep:US-08-467-822-26

seq_documentation_block:

Sequence 26, Application US/08467822

Patent No. 5843460

GENERAL INFORMATION:

APPLICANT: Labique, Agnes

APPLICANT: Sauerbaum, Sebastien

APPLICANT: Ferrero, Richard L.

APPLICANT: Thiberge, Jean-Michel

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

ADDRESSEE: Dunner

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-26

alignment_scores:
Quality: 92.50 Length: 424
Ratio: 0.517 Gaps: 22
Percent Similarity: 42.217 Percent Identity: 20.519

alignment_block:
US-09-303-518D-125 x US-08-467-822-26 ..
Align seg 1/1 to: US-08-467-822-26 from: 1 to: 569

205 ACTGCGCGCGCTTCAGCAAAATCGCGCGATTCACCGTGGCAAAAGC 254
|||||
161 ThrGlyProAlaAspGlyThrAsnAlaThrThrProGlyArg 177
255 CGTACTTACGTACGTGCGATTCGCGTGAAGCAACGACGAATCGAGT 304
|||
177 gasnleuylstrpmetleuargalala 186
305 TTGAACGCTACGACCTGAA.....GCGTGGCAAACTTAAGC 342
|||||
167 .GlucIlyrsermetasnleuglyPheleuAlaIalysglYAsnAlaSer 202
343 GCGCAAGAAGTGGCGCGACCTGATCCAAATCCGCTTG..... 381
|||||
203 AsnaSerAlaSerAlaArgAspGlnIleGluAlaGlyAlaIleGlyPhe 219
382 .....TGACTGCGCTGGCGACCGCTCCGCTTCAGCAAAATTC 418
|||||
219 slleHisgluAspTrp.....GlyThrThrProSerAlaIleAsn 233
419 CTGCGGTGCAT...GCGAGCGGTTCGCGCATCTTCATTCGATGCGAGC 465
|||||
233 lAlaIleuAspValAlaAspIysTyrAspValGlnVal...AlaIleHis 248
466 ACCAATCGCTGGCTGGCGACCGCTACGTCTATTCAAGAAGCC..... 510
|||||
249 ThrAspThrLeu.....AsnGluAlaGlyCys 257
511 .GCGGAGATTTCAAACGCGCGCTGTGTATTGAGCGTTTGACCGAAC 559
|||||
257 sValGluAsp.....ThrMetAlaIleAlaIleGly 268

```

```

560 GCAAAATCCATCTTTGTAAAGCAGCTGGCGGACGACGTGCGCTGAAAT 609
|||||
268 rGthrMetHisThrPheHisThrGluGlyAlaGly..... 279
610 GCTGCCAACATCGAAMACATGAAATTCGCGCGCCGATCCGCGTTT 659
|||||
280 .....GlyGlyHisAlaProAlaIle 287
660 GAGTGGCAGCGCACATTCATTCATCGAGCGGTGGCGCGAATTAACCG 709
|||||
287 elysValAlaGlyGluHisAsnIleLeuProAlaSerThrAsnProThr 304
710 TG...TGACCATCAAT.....TATCAAGATGTAATTACCAAT 744
|||||
304 lProPheThrValAsnThrGluAlaGluHisMetAspMetLeuMetVal 320
745 GCGCGT.....TTGTTTGAACAGG 764
321 CysHisHisLeuAspIysSerIleGlyGluAspValGlnPheAlaAsp 337
765 CGGTTCGAACACGAGCGCGTATGCC.....CTAG 796
|||||
337 rArgIleArgProGlnThrIleAlaAlaGluAspThrLeuHisAspMet 354
797 GTGGTTCTCAAGTCAACAAACCGCG..... 822
|||
354 lylPheSerIleThrSerAspSerGlnAlaMetGlyArgValGly 370
823 .....CTCTGCGTACGCTTTGGTGGCAAGATGCAAAATTAATCTGC 866
|||||
371 GluValIleThrArgThrTrpGlnThrAlaAspIysAsnLysGluPh 387
867 GGGCGAATTGTT.....GACACAGACAAAC.....C 892
|||||
387 eGlyArgLeuLysGluGluLysGlyAspAsnAspPheArgIleLys 404
893 GCGTGATTCGCGTTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 942
|||||
404 rGlyrLeuSerLysTyrThrIleAsnProAlaIleAlaHisIleSer 420
943 GATTATTTGGAGCTACACCAATCAATTCGTTATTCGATTCGATTCGAT 992
|||||
421 GluThrValGlySer.....ValGluValGly 430
993 CAGCAAGAGCTGTGCGCTG..... 1014
|||
430 sValAlaAspLeuValLeuTrpSerProAlaPheGlyValLysPro 447
1014 ..... 1014
447 smetIleIleLysGlyPheIleAlaLeuSerGlnMetGlyAspAla 463
1015 .....GTTCGCGCGGACCGGACAAATCTCCATCAGCGTAC 1052
|||||
464 AsnAlaSerIleProThrProGlnProValTyrTyr.....Arg 477
1053 AACCTGGGCGCAT.....TTCGGAAGAAACAACTCTTCA 1087
|||||
477 umetPheGlyHisHisIleLysAlaLysTyrAspArgAsnIleThrPhe 494
1088 AGTTCAACACACCGCTCAACGCGCGCGCGCCATGTCGCGATGCT 1137
|||||
494 alserGlnAlaIleTyrAspIysGlyLysGlu.....GluLeuGly 508
1138 ACTTACGAGCGCGTATGCTTCGATATCTGCGCACCGCTGCTTTGG 1187
|||||
509 leuGluArgGlnValLeuProValLysAsnGlyArgAsnIleThrLys 525
1188 CGATTATTCGTCGCGGATACC 1209
525 sAspMetGlnPheAsnAspThr 532

```


seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-432-697-26

seq_documentation_block:

Sequence 26, Application US/08432697
Patent No. 6248330

GENERAL INFORMATION:

APPLICANT: Ladique, Agnes
APPLICANT: Sauerdam, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-697-26

alignment_scores:

Quality: 92.50 Length: 424
Ratio: 0.517 Gaps: 22
Percent Similarity: 42.217 Percent Identity: 20.519

alignment_block:

US-09-303-518D-125 x US-08-432-697-26 ..

Align seg 1/1 to: US-08-432-697-26 from: 1 to: 569

```

205  ACTGGCGCGGCTTCAGCAAAATGCGCGCATTCACCGTGGCAAAAGCG 254
161  ThrgtlyProAlaaspGlyThrAsnAlaThrThrProGlyArgAr 177
255  CGTACTTCAGTACGTGATTCGCGTTCGTAAGCAAGCAAGCAATGAGT 304
177  gAsnleuIysrIrrpMetleuArgAlaAla..... 186
305  TTGAACGCTGCGACACTGAA.....GCGCTGGCAAACTTAAGC 342
187  ..GlugIuTySerMetAsnleuGlyPheleuAlaIysGlyAsnAlaSer 202
343  GGCAGAAAGTGGCGCGCAACTGATCCAAATCGGTTTG..... 381

```

```

203  AsnAspAlaSerAlaArgAspGlnIleGluAlaGlyAlaIleGlyPhe 219
382  .....TGGAOTGGCGCGCGCACCGCTCCGTCGCAAAATTC 418
219  sIleHisGluAspTrp.....GlyThrThrProSerAlaIleAsnH 233
419  CTGGCGCTGCAT...GCCGACCGCTGCGCAATTCGTCGCAATGGAGAC 465
233  IsAlaIleAspValAlaAspIlystYrAspAlaIleVal...AlaIleHis 248
466  ACCAATCCGCTGGCTGGCGGACCGCTACGGTCATGATTAAGAAAGCC..... 510
249  ThrSpThrleu.....AsnGluAlaGly 257
511  .GCCGAGATTCGAAACGGCGCTGTTGGATTCGACCCCTTGACCGAAC 559
257  sValGluAsp.....ThrMetAlaAlaIleAlaGly 268
560  GCAAAATTCATGTTTGTAAAGCAGCTGGCGCAGACGCTGGCTGAAAT 609
268  rGThrMetHisThrPheHisThrGluGlyAlaGly..... 279
610  GCTGCCAATCGAACAACATGATTCGGCGCGCGCATCTCGCGGTTT 659
280  .....GlyGlyHisAlaProAspIleI 287
660  GAGTGGCAGCAGCATTCATTCATGACCGCGTGGCGGCAATTAACG 709
287  elysValAlaGlyGlnHisAsnIleleuProAlaSerThrAsnProIrr 304
710  TG...TGACACATCAAT.....TATCAAGATGATATACAT 744
304  leProPheThrValAsnThrGluAlaGlnHisMetAspMetleuMetVal 320
745  GGCGCT.....TGTTTGGCAACAG 764
321  CysHisHisLeuAspIysSerIleIysGluAspValGlnPheAlaAsp 337
765  CGGTGGAACACCGAGCGGTGATTCG.....CTAG 796
337  rArgIleArgProGlnThrIleAlaIleAlaGlyAspThrLeuHisAsp 354
797  GTGGTTCGAAAGTCAACAACCGCGC..... 822
354  IyIlePheSerIleThrSerAspSerGlnAlaMetGlyArgValGly 370
823  .....CTCTGGCTACCGTTTGGGTGCAAAAGTATCGCAAAATTACTGC 866
371  GluValIleThrArgThrTrpGlnThrAlaAspIysAsnIysGluPhe 387
867  GGGCGAATTTGTT.....GACACAGACAC.....C 892
387  eGlyArgLeuIysGluGluIysGlyAspAsnAspAsnPheArgIleIys 404
893  GCGTGATTTCCGCTTCGATTCGATTCGACGCGCGGATTCACAGCGCG 942
404  rGlyrLeuSerIysTrpThrIleAsnProAlaIleAlaHisGlyIleSer 420
943  GATATTTGGACCGCTACCAATCAGATTCGCGTATTCGAGAAAGCGCG 992
421  GluIyValGlySer.....ValGluValGly 430
993  CAGCAAGAGCTGTTCGGCTG..... 1014
430  sValAlaAspIleValleuTrpSerProAlaPhePheGlyValIysPro 447
1014  ..... 1014
447  snMetIleIleIysGlyGlyPheIleAlaIleuSerGlnMetGlyAspAla 463
1015  .....GTGGCGCGCGCGCGGACAAATTCATCATACCGCTAC 1052
464  AsnAlaSerIleProThrProGlnProValIyTrp.....ArgI 477

```

```

1053 AACCTCGGCAT.....TTCGTAAACAACTCTCA 1087
      ::|||::
477 uetpgegylhshlslyslalylstyrarpargsnlethrrphev 494
      ::|||::
1088 AGTTCAACACAGCCGTCACAGCGCGGCATGTCGCATGCT 1137
      ::|||::
494 alserglnalaatyraplysglylelysglu.....gluueuly 508
      ::|||::
1138 ACTTACGAGCGCGTATGCGCTTGATATCCGCCACCTGCTTTGCG 1187
      ::|||::
509 leuglaarglnvalleuprovalllysasnycysargasnilethlysl 525
      ::|||::
1188 CGATTTAATCGTCGCGCATACC 1209
      ::|||::
525 sasmetglnpheasnasprthr 532
      ::|||::

```

```

seq_name: /cgn2_6/ptodata/1/laa/6b_COMB pep:us-08-466-248-26

```

```

seq_documentation_block:

```

```

; Sequence 26, Application US/08466248
; Patent No. 6258359

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-248-26

```

```

alignment_scores:

```

```

      Quality: 92.50      Length: 424
      Ratio: 0.517      Gaps: 22
Percent Similarity: 42.217      Percent Identity: 20.519

```

```

alignment_block:

```

```

US-09-303-518d-125 x US-08-466-248-26

```

```

Align seg 1/1 to: US-08-466-248-26 from: 1 to: 569

```

```

205 ACTGCGCGCGCTTCAGGCAAAATCGCCGCGATTCACCGTCGCGCAAAAGC 254
      |||::|||::
161 thnglyproalaspdltyhrasnalathrrlthrrprogllyarghr 177
      |||::|||::
255 CGTACTTCAGTCAGTCGTGATGCCGTGAAGGCAAGCAAGCAAAATCGAGT 304
      |||::|||::
177 gsnleulystrpmetleuargalala..... 186
      |||::|||::
305 TTGAACGTACGACACCTGAA.....CGCTGGCAAACTTAAAC 342
      |||::|||::
187 ..gluglutysermetasnleuglypheleulalylslyasnalaser 202
      |||::|||::
343 GCGGAAGAAGTGGCGCGCAACCTGATCCATCCGCTTG..... 381
      |||::|||::
203 Asnaspalaseralargaspelnileglualaglyalaleglypely 219
      |||::|||::
382 .....TGACTGCGCTGCGCACCCGCTGTCAGCAAAATTC 418
      |||::|||::
219 sllehlsgluasprtr.....glythrthrrproseralaleasn 233
      |||::|||::
419 CTGCGCTGCAT..GCCGACCGCTGCCCATCTGTCAGAACGATGAGC 465
      |||::|||::
233 lsalaleuaspyalalaasplytyraspvalglnval...Alallehlis 248
      |||::|||::
466 ACCAATCGCTGCGCTGCGCACCTTACGTCATTATCAAGAAGCC..... 510
      |||::|||::
249 Thrasprthleu.....Asnluualaglycy 257
      |||::|||::
511 .GCCGAGATTTCAAAAGCGCGCTGTCATTTGACCGCTTGACCGAAC 559
      |||::|||::
257 svalgluasp.....Thmetalaalalealagly 268
      |||::|||::
560 GCAAAATCCATTTTGTGAAGCGAGTCGGCGACAGCTGCCCTGTAAT 609
      |||::|||::
268 rylthmetlnsthrpnehlsthrcluglyalagly..... 279
      |||::|||::
610 GCTGCAACATCGAAACACATGAATTCGGCGCGCGCGCTGCCGCTT 659
      |||::|||::
280 .....Glyglynlsalaproaspillel 287
      |||::|||::
660 GAGTGGACAGCACATTTATTTATTCATCGACCGCGCGCGGATTAACCG 709
      |||::|||::
287 elysvalalaglygluhlsasnleleuprolasethrrasprrothrl 304
      |||::|||::
710 TG...TGACCATCAAT.....TATCAAGATGTAATTCACAT 744
      |||::|||::
304 leprophethvalasnthrglualaglylnlmetksrmetleuval 320
      |||::|||::
745 GGCGGT.....TTGTTTGCACACAG 764
      |||::|||::
321 Cynshlsleuasrlyserllelysgluaspyalglrphelalaaspe 337
      |||::|||::
765 CCGTGTGAACACGAGCGCGGTGATGCG.....CTAG 796
      |||::|||::
337 rarglleargproglnthrllealalaagluasprthleuhlsasrmet 354
      |||::|||::
797 GGGTTCAGTCACAAACACCGCGC..... 822
      |||::|||::
354 lylllepheserllethrrsererserserglnalamegllyargvalgly 370
      |||::|||::
823 .....CTCTTGCCTACCGCTTTTGGTGGCAAGATACGCAATTTACTGC 866
      |||::|||::

```

```

371 GluValIleThrArgThrTrpGlnThrAlaAspLysAsnLysGluPhe 387
867 GGGGGAATTGGT.....GACACAGACAAC.....C 892
387 eGlyArgLeuLysGluGlnLysGlyAspAsnAspAsnPheArgIleLysA 404
893 GCGTGAATTCGGCTGATTGAACGGCGGATTAACACAGCGCGCAC 942
404 rGlyrLeuSerLysTrpThrIleAsnProAlaIleHisGlyIleSer 420
943 GATTATTGGAGACCTACCAACAATGATTCCTGATCGAAGAGCGCG 992
421 GluTrpValGlySer.....ValGluValGlyLys 430
993 CAGCAAGAGCTGTTCGGCTG..... 1014
430 sValAlaAspLeuValLeuTrpSerProAlaPheGlyValLysProA 447
1014 ..... 1014
447 smMetIleIleLysGlyPheIleAlaLeuSerGlnMetGlyAspAla 463
1015 .....GTGCGCGGAGCGCGGACAATTACTCATACGCGTAC 1052
464 AsnAlaSerIleProThrProGlnProValTrpTyr.....ArgG 477
1053 AACCTCGCCCAT.....TTCCTGAAAACAAACTCTTCA 1087
477 umetPheGlyHisHisGlyLysAlaLysTrpAspArgAsnIleThrPheV 494
1088 ACTTCACACAGCCCTCAACGCGCGCACCGCCATGTCGCGATTGCT 1137
494 aIserGlnAlaAlaLysTrpLysGlyLysGlu.....GluLeuGly 508
1138 ACTTACGAGCGGTGATGCTTCGATATCTCTGCCACCGTGGTTGCG 1187
509 LeuGlnArgGlnValLeuProValLysAsnLysArgAsnIleThrLys 525
1188 GATTATATCGTCGCGCATAC 1209
525 sAspMetGlnPheAsnAspThr 532

seq_name: /cgn2_6/prodata/1/1aa/B_B_COMB.pep:US-08-669-785-4
seq_documentation_block:
; Sequence 4, Application US/08669785
; Patent No. 6309648
; GENERAL INFORMATION:
; APPLICANT: Betsou, Fotini
; APPLICANT: Sebou, Peter
; TITLE OF INVENTION: Protective Epitopes of Adenyl
; TITLE OF INVENTION: Cyclase-Haemolysin(AC-Hly), Their Application To
; TITLE OF INVENTION: The Treatment Or To The Prevention Of Bordetella
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.

```

```

; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356,0072-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1705 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-669-785-4

```

```

alignment_scores:
Quality: 92.50 Length: 467
Ratio: 0.467 Gaps: 24
Percent Similarity: 42.398 Percent Identity: 21.413

```

alignment_block:

```
US-09-303-518d-125 x US-08-669-785-4 ..
```

```
Align seg 1/1 to: US-08-669-785-4 from: 1 to: 1705
```

```

67 GCGCCGCCATTACCGAAGTCGCTGCTTGGCGAAGATATGCGCGTAT 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
773 GlyValGlnThrThrGluIleSerLysSerAlaLeuGluLeuAlaIle 789
117 GCGCCGCCATTACCGAAGTCGCTGCTTGGCGAAGATATGCGCGTAT 166
: : : : : : : : : : : : : : : : : : : : : : : : :
789 e.....ThrGlyAsnAlaAspAsnLeuLysSerAlaAspY 801
167 TCGTGTGAAGACAAAAGATCCGGC..... 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 a.....PheValAspArgPheIleGlnGlyLysArgValAlaGlyGlnPro 816
196 GTGGGTTCCTGCGCGCGCTTCAAGCAAAATGCGCGCATTCACCGTGG 245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
817 ValValLeuAspValAlaAlaGlyGlyIleAspIleAlaSerArgLys 833
246 CGAAAGCGCGTACTTCAGTCAGTCGATTCGCGTTCGTTGAAGGCAACGACG 295
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
833 YGluArgProAlaLeuThrPheIleThrProLeuAla..... 845
296 AATCGAGTTGAACGCTACGACCTGAAGCGCTGCAACTTAACCGCG 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
846 .....AlaPro.....Gly 848
346 GAAGAGTGGCGCGCAACCTGATCAATCCGTTGTGACTGCGCTGG 395
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 GluGluGlnArgY..... 854
396 CACCGCGCTTCAAGCAAAATTCCTGCGTTCGATTCGCGCGATTCGCGCA 445
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 gThrLys.....ThrGlyLysSerGlnPheThrT 864
446 TCTTCGCAATGCGATGAGACCAATCCGCTG..... 477
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
864 hrPheValGluIleValGlyLysGlnAspArgTrpArgIleArgAspGly 880
478 GCTGCGGACCTACGCTCATTAACAAGAGCGCGGAGATTCGAACG 527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
881 AlaAlaAspThrThrIleAspLeuAlaLys..... 890
528 CGCGCTGTTGATGAGCGCTTGAACGCAACG..... 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
891 .....ValValSerGlnLeuValAspAlaAsnGlyValLeuLys 904
562 .....AAATCGATGTTGTGAAGCGAGCTGCGCGAGATTCGCGCT 603
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904 IserIleLysLeuGlnValIleGlyLysAspGlyAspAspValValLeu 920

```



```

126 .....ProIleGlyHisIleuValaIspAlaIleuArgIleGly 138
127 .....
128 .....
129 TCGAGTTTGAACGCTACGACCTGAGCGGCGCAAACTTAAGCGCGAA 348
130 .....
131 .....
132 .....
133 .....
134 eugIleuGlnGlnAsnTyr.....AspTyrL 144
135 .....
136 .....
137 .....
138 .....
139 GAAAGCGCGCCCAACCTGATCCAAATCCGTTTGTGACTGCGTGGCAG 398
140 .....
141 .....
142 .....
143 .....
144 .....
145 .....
146 .....
147 .....
148 .....
149 .....
150 .....
151 .....
152 .....
153 .....
154 .....
155 .....
156 .....
157 .....
158 .....
159 .....
160 .....
161 .....
162 .....
163 .....
164 .....
165 .....
166 .....
167 .....
168 .....
169 .....
170 .....
171 .....
172 .....
173 .....
174 .....
175 .....
176 .....
177 .....
178 .....
179 .....
180 .....
181 .....
182 .....
183 .....
184 .....
185 .....
186 .....
187 .....
188 .....
189 .....
190 .....
191 .....
192 .....
193 .....
194 .....
195 .....
196 .....
197 .....
198 .....
199 .....
200 .....
201 .....
202 .....
203 .....
204 .....
205 .....
206 .....
207 .....
208 .....
209 .....
210 .....
211 .....
212 .....
213 .....
214 .....
215 .....
216 .....
217 .....
218 .....
219 .....
220 .....
221 .....
222 .....
223 .....
224 .....
225 .....
226 .....
227 .....
228 .....
229 .....
230 .....
231 .....
232 .....
233 .....
234 .....
235 .....
236 .....
237 .....
238 .....
239 .....
240 .....
241 .....
242 .....
243 .....
244 .....
245 .....
246 .....
247 .....
248 .....
249 .....
250 .....
251 .....
252 .....
253 .....
254 .....
255 .....
256 .....
257 .....
258 .....
259 .....
260 .....
261 .....
262 .....
263 .....
264 .....
265 .....
266 .....
267 .....
268 .....
269 .....
270 .....
271 .....
272 .....
273 .....
274 .....
275 .....
276 .....
277 .....
278 .....
279 .....
280 .....
281 .....
282 .....
283 .....
284 .....
285 .....
286 .....
287 .....
288 .....
289 .....
290 .....
291 .....
292 .....
293 .....
294 .....
295 .....
296 .....
297 .....
298 .....
299 .....
300 .....
301 .....
302 .....
303 .....
304 .....
305 .....
306 .....
307 .....
308 .....
309 .....
310 .....
311 .....
312 .....
313 .....
314 .....
315 .....
316 .....
317 .....
318 .....
319 .....
320 .....
321 .....
322 .....
323 .....
324 .....
325 .....
326 .....
327 .....
328 .....
329 .....
330 .....
331 .....
332 .....
333 .....
334 .....
335 .....
336 .....
337 .....
338 .....
339 .....
340 .....
341 .....
342 .....
343 .....
344 .....
345 .....
346 .....
347 .....
348 .....
349 .....
350 .....
351 .....
352 .....
353 .....
354 .....
355 .....
356 .....
357 .....
358 .....
359 .....
360 .....
361 .....
362 .....
363 .....
364 .....
365 .....
366 .....
367 .....
368 .....
369 .....
370 .....
371 .....
372 .....
373 .....
374 .....
375 .....
376 .....
377 .....
378 .....
379 .....
380 .....
381 .....
382 .....
383 .....
384 .....
385 .....
386 .....
387 .....
388 .....
389 .....
390 .....
391 .....
392 .....
393 .....
394 .....
395 .....
396 .....
397 .....
398 .....
399 .....
400 .....
401 .....
402 .....
403 .....
404 .....
405 .....
406 .....
407 .....
408 .....
409 .....
410 .....
411 .....
412 .....
413 .....
414 .....
415 .....
416 .....
417 .....
418 .....
419 .....
420 .....
421 .....
422 .....
423 .....
424 .....
425 .....
426 .....
427 .....
428 .....
429 .....
430 .....
431 .....
432 .....
433 .....
434 .....
435 .....
436 .....
437 .....
438 .....
439 .....
440 .....
441 .....
442 .....
443 .....
444 .....
445 .....
446 .....
447 .....
448 .....
449 .....
450 .....
451 .....
452 .....
453 .....
454 .....
455 .....
456 .....
457 .....
458 .....
459 .....
460 .....
461 .....
462 .....
463 .....
464 .....
465 .....
466 .....
467 .....
468 .....
469 .....
470 .....
471 .....
472 .....
473 .....
474 .....
475 .....
476 .....
477 .....
478 .....
479 .....
480 .....
481 .....
482 .....
483 .....
484 .....
485 .....
486 .....
487 .....
488 .....
489 .....
490 .....
491 .....
492 .....
493 .....
494 .....
495 .....
496 .....
497 .....
498 .....
499 .....
500 .....

```

```

350 Thr.....GluLeuArgIleValaIle 358
351 CGTCACGCGCGCGACCGCCCATGATGCGCAATGTACTACAGCGCG 1150
352 .....
353 .....
354 .....
355 .....
356 .....
357 .....
358 .....
359 .....
360 .....
361 .....
362 .....
363 .....
364 .....
365 .....
366 .....
367 .....
368 .....
369 .....
370 .....
371 .....
372 .....
373 .....
374 .....
375 .....
376 .....
377 .....
378 .....
379 .....
380 .....
381 .....
382 .....
383 .....
384 .....
385 .....
386 .....
387 .....
388 .....
389 .....
390 .....
391 .....
392 .....
393 .....
394 .....
395 .....
396 .....
397 .....
398 .....
399 .....
400 .....
401 .....
402 .....
403 .....
404 .....
405 .....
406 .....
407 .....
408 .....
409 .....
410 .....
411 .....
412 .....
413 .....
414 .....
415 .....
416 .....
417 .....
418 .....
419 .....
420 .....
421 .....
422 .....
423 .....
424 .....
425 .....
426 .....
427 .....
428 .....
429 .....
430 .....
431 .....
432 .....
433 .....
434 .....
435 .....
436 .....
437 .....
438 .....
439 .....
440 .....
441 .....
442 .....
443 .....
444 .....
445 .....
446 .....
447 .....
448 .....
449 .....
450 .....
451 .....
452 .....
453 .....
454 .....
455 .....
456 .....
457 .....
458 .....
459 .....
460 .....
461 .....
462 .....
463 .....
464 .....
465 .....
466 .....
467 .....
468 .....
469 .....
470 .....
471 .....
472 .....
473 .....
474 .....
475 .....
476 .....
477 .....
478 .....
479 .....
480 .....
481 .....
482 .....
483 .....
484 .....
485 .....
486 .....
487 .....
488 .....
489 .....
490 .....
491 .....
492 .....
493 .....
494 .....
495 .....
496 .....
497 .....
498 .....
499 .....
500 .....

```

```

seq_name: /cgn2.6/plodata/1/1aa/5A_COMB.pep:US-08-306-063-60
seq_documentation_block:
; Sequence 60, Application US/08306063
; Patent No. 5633435
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: Glycosylase Tolerant
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dennis R. Hoerner, Jr., Monsanto Co. B4F
; STREET: 700 Chesterfield Village Parkway
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,063
; FILING DATE: 13-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,611
; FILING DATE: 28-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/576,537
; FILING DATE: 31-AUG-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoerner Jr., Dennis R.
; REGISTRATION NUMBER: 30,914
; REFERENCE/DOCKET NUMBER: 38-21(10660)A
; TELEPHONE: (314)537-6099
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 60:
; LENGTH: 427 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-306-063-60

```

```

alignment_scores:
Quality: 92.00 Length: 431

```

Ratio: 0.474 Gaps: 21
Percent Similarity: 45.012 Percent Identity: 20.882

alignment block:
US-09-303-518d-125 x US-08-306-063-60 ..

Align seg 1/1 to: US-08-306-063-60 from: 1 to: 427

```

67 GGGCGGGCCATACCGAAGTCCGCTG.....CTTGGCGAAGATA 107
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
81 GLYLSLeuValAlaGluInProLeuGluLeuPheLeuGlyAsnAlaG1 97
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
108 TGGCGGTATGCGCCG.....TCGATGAAGTCAAGGAAGCGCATG 148
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
97 yThrAlaMetArgProLeuAlaAlaLeuGlyLeuGlyAsnAsp1 114
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
149 CCCTCAAAAAGGCCAAGTCTGTTGAAGACAAAAGATCCGGCGCTG 198
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
114 LeValLeuThrGlyGluProArgMetLysGluArg..... 125
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
199 GTGTTACTGCGCGGCTTCAGGCAAAATCGCGGATTCACCGTGGCGA 248
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
126 .....ProIleGlyHisLeuValAspAlaLeuArgGlnGly1 138
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
249 AAAGCCGTACTTCAGTCAAGTCTGATTCGCGTTGAAGCAACGCGAAA 298
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
138 yAlaGlnIle.....AspTyrFL 144
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
299 TCGAGTTGAAGCTACGCGACCTGAAGCGCTGGCAAACTTAAGCGCGAA 348
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
144 euGluGlnGluAsnTyr..... 149
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
349 GAAATGCGCGCAACTGATTCATCCGTTGTGACTGCGCGCGAC 398
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
150 .....ArgArgCysIleAlaGlyGlyPheArgGlyGlyLeuThr 163
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
399 CCGTCCGTCAAGCAAAATTCCTGCGCTGATGCGGAGCGCTTCGCATCT 448
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
163 T.....ValAspGlySerValSerGlnP 172
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
449 TCGTCAATGCG...ATGACACCAATCCGCTGCGCCACCTACGGTC 495
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
172 heLeuThrAlaLeuLeuMetThrAlaProLeuAlaGluGlnAspThrGlu 188
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
496 ATATCAAAAGAGCGCGCGAGGATTCAAAGCGCGCTGTGATTGAG 545
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
189 IleGlnIleGln.....GlyLeuLeuValSerIy 198
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
546 CCGTTTGAACGCAAGCAAAATCCATGTTGTGAAGCGACTGCGCGACAG 595
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 sProtyrIleAspIleThrIleuHisLeuMetLysAlaPheGlyValAsp 215
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
596 TGCCCTCTGAAAATCCCTGCCAAATCGAAACACATGAATTCGGGCGCGC 645
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
215 aValHisGluAsnTyrGlnIlePhe.....HisIleGlyGly... 228
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
646 CATCCTGCGCGTTGTGATGCGACGACCATATTCATTCAGCGCGGTGCG 695
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
229 .....GlnThrTyrArgSerProGlyIleTyrLeuValGln..... 240
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
696 CCGCAATAAACCGTGTGACCATCAATATCAAGATGATTAATCCATG 745
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
241 .....GlyAspAlaSerSerIas 247
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
746 GCGGTTGTGTTGCAACAGCGCGCTG.....AACACGAGCGCGGTGAT 789
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
247 eTyrPheLeuAlaAlaAlaIleLysGlyThrValArgValThr 263
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
790 GCGCTAGTGTGTTCAAGTCAACAAACGCGCGCTTCGCTACCGTTT 839
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
264 GlyIleGlyLysGlnSerValGlnIleLysPThrLysPheAlaAspVal 280
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
840 G.....GGTGGCAAGTATCG.....CAATTA 862

```

```

280 uGluLysMetCylValAlaLysIleSerTyrPGLysAspTyrIleGluCys 297
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
863 GTGCGGCGCAATTC.....GTGACACAGACAAACCGCGATTCGCCG 906
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
297 eArgGlyGluLeuGlnGlyIleAspMetAspMetAsnHisIleProAsp 313
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
907 TCGGATTTGAACGGCGCATTCACAAAGCGCGCGCATTAATTTGGAGC 956
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
314 AlaAlaMetThrIleAlaThrThr..... 321
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
957 CTACCAACATGATGATTCGTTATTCGAAGAAGCGCGAC.....AAG 1000
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
322 .....AlaLeuPheAlaAspGlyProThrValIleArg 333
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1001 ACCTGTTCGCGCTGGTTCGCGCGCGACCGCGCAAAATCTCCATCAGCG 1050
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
333 snIleTyrAsnTyrParValLysGluThrAspArgLysLeuSerAlaMetAla 349
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1051 ACACCCCTGCGCATTTCTCTGAAAAAACAACCTTCATCAATTCAACACAG 1100
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
350 Thr.....GluLeuArgLysValGlyAlaG1 358
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1101 CGTCAAGCGCGCGACCGCGCGCATGTGCGCATTCATTCAGAGCGCG 1150
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
358 uValGluGluGlyGln.....AspTyrIleArg 368
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1151 TGATGCGCTTGATATCTGCGCCACCTGCTTTGCGCGATTAATC... 1197
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
368 aValAlaPro.....ProAlaGlnLeuIleAlaGluIleGly 380
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1198 ...GTGCGGATACGACGACCGCGCGCATTCGCTTCGGAATTCGA 1244
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
381 ThrTyrAsnAspHisArgMetAlaMetCysPheSerLeuValAlaLeuSe 397
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1245 CGAAGAAGACCTGCTTGTGCGAGCTTCGCTGCGCGGCAAA 1287
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
397 rAspThrProValThrIleLeuAspProLysCysThrAlaLys 411
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /cgn2.6/ptodata/1/1aa/5A_COMB.dep:US-08-833-485-60

```

seq documentation block:
Sequence 60, Application US/08833485
Patent No 5804425
GENERAL INFORMATION:
APPLICANT: Barry, Gerard F.
APPLICANT: Kishore, Ganesh M.
APPLICANT: Padgett, Stephen R.
APPLICANT: Stallings, William C.
TITLE OF INVENTION: Glycosylate tolerant
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
STREET: 700 Chesterfield Village Parkway
City: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,485
FILING DATE: 07-Apr-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,063
FILING DATE: 13-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

546  CGCTTTGACGACGACGACAAATTCATGTTTATTAAAGCAAGCTGGCCGACAGC  595
      :               : : : : : : : : : : : : : : : : : : : : : :
198  sProTyrlIleAspIleThrIleuHisIleuMetLysAlaPheIleValAspY  215
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
596  TGGCGTGTGAAATGCTGCCAATCGAATGAAACACATGATGAAATGGCGGGCCG  645
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215  aIValHisIleuInsuTyrgIleIlePhe.....HisIleLysGlyGly...  228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
646  CATCCGCGCGGTTTGAGTGGCAGCGACATTCATTTTCATCGAGCGGGTCGG  695
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
229  .....GlnThrTyArgSerProGlyIleTyLeuValGlu.....  240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
696  CGCGAATAAAGCCGTGGACCATCAATTAATCAAGATGTAATTAACATTG  745
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241  .....GlyAspAlaSerSerAla  247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
746  GCCGTTTGTTCACACAGCGCGTCTG.....AACACGAGCGCGTGAAT  789
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247  eTyThrPheLeuAlaIleAlaIleValIleLysGlyIleThrValArgValThr  263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
790  GCCCTAGAGGTTCACAGTCAACAAACACGCGGCGCTTCCTGTAACGCTTT  839
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264  GlyIleGlyLysGlnSerValGlnIleLysPheThrLysPheAlaAspValLe  280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
840  G.....GGTGGCAAGTATCG.....CAATTA  862
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
280  uGluLysMetGlyAlaLysIleSerThrGlyAspAspTyrlIleGluLysS  297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
863  CTGGCGGCGAATTCG.....GTTCACACAGACACACCGCGTATTCGGGT  906
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
297  eArgGlyGlnIleuGlnIleLysPheLysPheLysHisIleProAsp  313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
907  TCGGATTTGAACGCGCGCATTCACACAGCGCGACGATTAATTTGGAGC  956
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
314  AlaAlaMetThrIleAlaThrThr.....  321
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
957  CTACCAACATAGATTTCCGTTATCGAAGAGCGCGAC.....AAAG  1000C
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
322  .....AlaLeuPheAlaAspGlyProThrValIleArgA  333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1001  AGCTGTTCGCGTGGGTGGCGGCGGACGCGGACCAAAATCTCCATCAGCGGT  1055C
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
333  snIleTyArgIleThrPArgValLysGluThrAspArgLeuSerAlaMetAla  349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1051  ACAACCTCGGCGCATTCCTCGAAAACAAACTCTTCAAGTTCAACACAGC  1100C
      || ||||| ||||| ||||| ||||| ||||| ||||| |||||
350  Thr.....GluLeuArgLysValIleGlyAlaG  358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1101  CGTCAAGCGGCGGACCGCGGCATGTGCGCATTTGTACTTACGAGCGC  1150C
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358  uValGluGluGlyGln.....AspTyrlIleArgY  368
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1151  TGAATGCCCTGGATATCTGCCACACCTCGCTTTTGCGCGAATTAATC...  1197
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
368  aIValPro.....ProAlaGlnIleuIleAlaIleGluIleGly  380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1198  ...GTGGCGCATTCGACGACGCGGACGCAAGCAATGGGTTGCTTGAATTGA  1244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
381  ThrTyArgAsnHisArgMetAlaMetCysPheSerIleuValAlaIleuSe  397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1245  CGAAGAAAGACCTCGCTTGTGACAGCTTCGTCCTCCCGGCGAA  1287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
397  RasPThrProValThrIleLeuAspProLysCysThrAlaLys  411
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-137-440-60

seq_documentation_block:
; Sequence 60, Application US/09137440
; Patent No. 6248876
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.

```



```

815 AACCGCGCCTCTGGCGA..... 832
277 ArglystyrlnsProthrlnsnlleuaspilleysglnlyProlysgl 2933
      ::::: :::::
833 .CCGTTT.....TGGGTGGAAAGTATGCAAAATTCTGGCGGCG... 871711
      ||||| ::::: ||| |||||
293 uProPheglnserrtyValasparghetylysserleuarglnag 3100
      .....
872 .....AATTGGTTGACACAGACACCGCGCTGATT 901
310 lntlrpspproAlaValylvsasntpmethrlnThrLeu..... 3244
902 CCGGTTGCGATTGAACGGCGGATTATACAAAGGCGCAGATTATTG 951
325 .....lleglnsnAlasnPro..... 330
952 GGAAGCTACACACATACATTCCGTTATCGAAGAAGCGCGACAGAAGA 100
331 .AspCysLysLeuValLeuLysglYleuGly.MetasnProthrLeuGln 346
1002 GCATTGCGCGTGGGTGGCGCCGACCGCGA..... 103
347 GluMetLeuThrAlaCysglnlyValglYlelyProolYglnlysnAlar 363
1032 .....CAAAATTCATCACGCGGTACAAACCTCGGCCATTTCCT. 1074
363 glauMetAlaGlnAlaLeuLysGlnAlaMetGlyProserProIleProp 380
1071 .....GAAAACAAACTCTTCACAGTCAACGC..... 109
380 heAlaIaAlaGlnlnrArglysnAlaIleArglytTrpsnCySgLyLys 396
1098 .....AGCCGTCAACGGCGGCGACCGCGCCATGTCGCCGATTG 1135
397 glnglYlnsSerzAlaArglncysArgAlaProArgArglnelYcstr 413
1136 GTACTTACAGAGCGCGTGATGCCCTTGGATATCCTGGCCACCGCTGTTTG 1185
413 P.....LysCysgLyLysProglYlns...lIemeta 423
1186 CCGCATTTAATCGTCGGCGGATACGACAGCGCGCAGCATGGAGTTGGTCT 1235
423 lAsnCyspRoglUargln.....AlaelyPheleu 433
1236 GGAATTGGACGAAGA.....AGACCTCGC 1259
434 GlyLeuGlyProArgGlyLysProArg 443

seq_name: /cgn2_6/prodata/1/1aa/5A_COMB.pep:US-08-427-097-2

seq_documentation_block:
? Sequence 2, Application US/08427097
? Patent No. 5668294
? GENERAL INFORMATION:
? APPLICANT: Meagher, Richard B.
? APPLICANT: Sommers, Anne O.
? TITLE OF INVENTION: Metal Resistance Sequences and
? TITLE OF INVENTION: Transgenic Plants
? NUMBER OF SEQUENCES: 30
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Greentree, Winner and Sullivan, P.C.
? STREET: 5370 Manhattan Circle, Suite 201
? CITY: Boulder
? STATE: CO
? COUNTRY: US
? ZIP: 80303
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentln Data #1.0, Version #1.30
? CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/427,097
 FILING DATE: 21-APR-1995
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Feider, Donna M.
 REGISTRATION NUMBER: 33,878
 REFERENCE/DOCKET NUMBER: 40-94
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 499-8080
 TELEFAX: (303) 499-8089
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 564 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-427-097-2

alignment_scores:
 Quality: 92.00 Length: 421
 Ratio: 0.479 Gaps: 17
 Percent Similarity: 45.606 Percent Identity: 20.190

alignment_block:
 US-09-303-518D-125/rev x US-08-427-097-2 ..

Align seq 1/1 to: US-08-427-097-2 from: 1 to: 564

```

1122 GGGCGGTCGCGCGCGCTTGACGGCTGTGTAAGTGAAGAGTTGTTT 1073
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 GlyIyIleAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 179
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1072 TCAGGAATGCGC.....CAGGCTTGACGCGCGATGAG 1038
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 aGInGInGInAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 196
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1037 TATTTGTCGCGCGCGCGACCCGACGACGCTTT.....GCT 994
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 Le.....LeuGIuGIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 209
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
993 GCGGCTTCCTTCGATTAACGGAATCTGATTTGCTA.....GC 956
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 226
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
955 GTCCCAATATATCGTGGCGCGCTTGTAATCGCGC..... 919
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 yGIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 243
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
918 .....GTTCAATACGGAACCGGA 901
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 roAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 259
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
900 AATCAGCGGCTGTCTGT..... 883
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 ThrGIuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 276
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
883 ..... 883
276 ySerSerValIeuaIaIeuaIaIeuaIaIaIaIaIaIaIaIaIaIaIaIa 293
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
882 .....GTCAACCAATTCGCGCGAGTAATTCGATCTTCGACCCAA 838
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 IaIyValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 308
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
837 AACGCTACCAAGAGCG.....CGGTTGTTGA 809
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 ProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 CTTGAGAACCACTAGGCAATCAGCGCTGCTTCAGACGCGCTGTT 759
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 uValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 338
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

758 GCAACCAACGCGCGAGTATTAATTCATCTGATTAATGATGTCGCAAC 709
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 IaasnGIuGIuGIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 352
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 GGTATTAATTCGCGCGCGAGCGCGCTGCATGAATGAATGATGCGTCA 659
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 GIyGIuIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 369
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
658 AACCGGCAAG.....ATCGGCGCGCGC 636
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 nThIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 386
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 AATCATGTGTTCGATGTGCG.....ACATTTTCAGACGCGAC 595
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 IaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 402
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 GTCTGCGCGAGCGCTTACAAACATGATTTGCGTTCGCTGCGTCAACG 545
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 415
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 TCATATCAACCAAGCGCGCTTGAATCCGCGCGCTTTGATATAG 495
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 .....ValAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 428
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
494 ACCGTAGGCTCGCGACCGCGATTCGTCG..... 462
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 ThrGIyGIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 445
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461 .....ATCGATTCAGCAAGATGCGCAACGCGCTCGGATCGA 425
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
445 eThIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 462
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 CGCGAGCAATTTGCTGAACGAGCGGTCGCGACGCGACGTCACAAACG 375
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 IaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 478
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 GATTCGATTCAGGTCGCGCGCGCACTTCGCGCGCTTAAGTTGCGCAG 325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 ArgAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 495
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 TTCAGTGCCTAGCGCTTCAACTGATTCGTCGCTTCCTTAACGCGCA 275
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 uGIuGIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 505
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 TCACGACTGACTGACGATGACGCGCTTTCGCGCGGATTCGCGGAT 225
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 .....AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 507
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 TTGCTGAAGCGCGCGAGTAACACCGCGCGATTCCTTTGCTTC 175
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 524
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 AACAGCACTTGCGCTTTTTCAGCGCATC.....CCTCTCTGACTT 131
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 nArgMeThIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 541
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 TCATCGAGCGG 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 eIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 544
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-427-097-14
 seq_documentation_block:
 ; Sequence 14, Application US/08427097
 ; Patent No. 566294
 ; GENERAL INFORMATION:
 ; APPLICANT: Meagher, Richard B.
 ; APPLICANT: Sommers, Anne O.
 ; TITLE OF INVENTION: Transgenic Plants
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee, winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/427,097
FILING DATE: 21-APR-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 40-94
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8080
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-427-097-14

Alignment scores:
Quality: 92.00 Length: 421
Ratio: 0.479 Gaps: 17
Percent Similarity: 45.606 Percent Identity: 20.190

alignment_block:

US-09-303-518D-125/rev x US-08-427-097-14

Align seg 1/1 to: US-08-427-097-14 from: 1 to: 564

1122 GCGCGCGTCCGCGCGTGTGACGCGCTGTGAACTTGAAGTGTGTTT 1073
163 GlyGlyLeuAlaLeuThrProThrLeuAlaLeuLeuAla 179
1072 TCAGGAATGCGC.....GAGGTGTACGCGGTGAG 1038
179 aglnnglnAlaArgValAspLeuArgHisAlaLysTyrGlnGly 196
1037 TATTTCGCGCGTCCGCGCGCAACCGCGACAGCTCTT.....GCT 994
196 Le.....LeuGlnGlyAsnProAlaLeuThrValLeuHisGlySer 209
993 GCGGCTTCTTCGATACGGAATGTGATGTGGA.....GC 956
210 AlaArgPheLysAspAsnArgAsnLeuLeuAlaGlnLeuAsnArgGly 226
955 GTCCCAATATATCTGCGCGCGCTGTGTAATCGCGC..... 919
226 ygluArgValAlaAlaPheAspArgCysLeuLeuAlaThrGlyAlaSer 243
918GTTCATACCGCAACCGCA 901
243 roAlaValProProLeuProGlyLeuLysAspThrProTyrThrSer 259
900 AATCAGCGCGTGTGCTGT..... 883
260 ThrGlnAlaLeuValSerGlnThrLeuProLysArgLeuAlaValIleG 276
883 883
276 ySerSerValAlaLeuGlnLeuAlaGlnAlaPheAlaArgLeuGly 293
882GTCAACCAATTCGCGCGCGATATTGGATGATCTTCACACCAA 838

293 lAlaValThrLeuAlaArgSerThrLeu...PhePheArgGluAsp 308
837 AACGTACGCAAGCGC.....CGTTTGTGA 809
309 ProAlaLeuGlnAlaValThrAlaAlaPheArgMetLeuGlyIleG 325
808 CTGAGACACCACTAGGCAATCAGCGCGCTCGGTTCAGACGCGCTGT 759
325 uValArg.....GluHisThrGlnAlaSerGlnValAlaTyr 338
758 GCAACAACGCGCAATGTAATTCATCTGTAAATGATGTCACAC 709
338 leasnGlyGlnGlyAspGlyGlnPheValLeuThr.....AlaHis 352
708 GGTTTTATTCGCGCGCGCGCTCGATGAAATGATGTCACAC 659
353 GlyLeuLeuArgAlaAspLysLeuLeuAlaThrGlyArgAlaProAs 369
658 AACGCGCAGG.....ATGCGGCGCGCGC 636
369 nThrArgLysLeuAlaLeuAspAlaThrGlyValThrLeuThrProGln 386
635 AATTCATGCTTCGATGTGCGC.....AGCATTTTCAGACGCGAC 595
386 LysAlaLeuValIleAspProGlyMetArgThrSerValGlnHisIleTyr 402
594 GTCGCGCGCAGTCCCTTACAACATGATTTGGCTGCGTCACACGCG 545
403 AlaAlaGlyAspCysThrAspGlnProGlnPheValTyr..... 415
544 TCAATACCAACAGCGCGCTTTGAAATCCTGCGCGCTTTGTAATG 495
416ValAlaAlaAlaAlaGlyThrArgAlaAlaIleAsnMet 428
494 ACCGTAGGCGCGCGCAGCGATTTGGGTCGCGCGCGCGCGCGCG 462
429 ThrGlyGlyAspAlaAlaLeuAsnLeuThrAlaMetProAlaValAlaPh 445
461ATCGCATTCAGCAAGATGCGCAACGCGCTCGCGCATGCA 425
445 eThrAspProGlnValAlaThrValGlyTyrSerGlnAlaGlnAlaHis 462
424 CGCAGCAATTTTGGTGAAGGAGCGGTGCGCGCGCGCGCGCGCGCG 375
462 IsaspGlyLeuLysThrAspSerArgThrLeuThrLeuAspAsnValPro 478
374 GATTGATCAGGTTGCGCGCGCACTTCTTCGCGCGCTTAAGTTGCGCAGCGC 325
479 ArgAlaLeuAlaAsnPheAspThrArgLysPheIleLysLeuValIleG 495
324 TTCAGGTGCGGTGCGCTTCAAACTCGATTTGCTGCTTCGCTTCAACGCGCA 275
495 uGlnLysSerGlyArgLeuIleGlyValGln..... 505
274 TCAGCACTGACTGAAGTACGCGCTTTTCGCGCGGTGATCGCGCGCAT 225
506AlaVal 507
224 TTGCTGAAAGCGCGCGCGATTAACACACGCGCGATTTTGTGCTTC 175
508 AlaProGlnAlaGlyGlnLeuLeuIleGlnThrAlaAlaLeuAlaIleArg 524
174 AAACAGCACTTGGCTTTTGAAGCGCATG.....CCTTCCTTACTT 131
524 nArgMetThrValGlnGlnLeuAlaAspGlnLeuPheProTyrLeuThr 541
130 TCATCGAGCGG 120
541 etValGlnGly 544

seq_name: /cgn2_6/ptodata/1/laa/5A_COMP.pep:US-08-427-097-16

```

seq documentation block:
; Sequence 16, Application US/08427097
; Patent No. 5668294
; GENERAL INFORMATION:
; APPLICANT: Meagher, Richard B.
; APPLICANT: Sommers, Anne O.
; TITLE OF INVENTION: Metal Resistance Sequences and
; TITLE OF INVENTION: Transgenic Plants
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/427,097
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Feher, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 40-94
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-427-097-16

alignment_scores:
      Quality: 92.00      Length: 421
      Ratio: 0.479      Gaps: 17
      Percent Similarity: 45.606      Percent Identity: 20.190

alignment_block:
US-09-303-518d-125/rev x US-08-427-097-16 ..
Align seg 1/1 to: US-08-427-097-16 from: 1 to: 564

1122 GCGCGGTCGCGCGCGCGTGTGACGCGTGTGAACTTGAAGATTGTTT 1073
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 GYGLYLLAIALAIALAIALAIALAIALAIALAIALAIALAIALAIALA 179
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1072 TCAGCAATGAGCC.....GAGGTTGTACGCGTGTATGAG 1038
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 agtnglgnlgnlgnlgnlgnlgnlgnlgnlgnlgnlgnlgnlgnlgnl 196
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1037 TATTTGTCCGCGTCCGCGCAACCGGCAACAGCTCTT.....GCT 994
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 le.....leuglgnlgnlgnlgnlgnlgnlgnlgnlgnlgnlgnlgnl 209
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
993 GCGCGCTTTCGATACGCAATGATGATGTA.....GC 956
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 AlAaRphelYsPaSnArGySlnleuIleValGlnleuAsnAspGlyGI 226
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
955 GTCCCAATATATGTCGCGCGCTTGTATATCGCGC..... 919
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 YGLuARgValIleValAlaPheAspArGySlnleuIleAlaThrGlyAlaSerP 243
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
918 .....GTTCAATACCGAACCGGA 901
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

243 roAlaValProIleProGlyLeuLysAspThrProTyrrTrpThrSer 259
900 AATCAGCGGTTGTCTGT..... 883
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 ThrGluAlaLeuValSerGlnThrIleProLysArgLeuAlaValIleGI 276
883 ..... 883
276 YSerSerValLeuAlaLeuGlnLeuAlaGlnAlaPheAlaArgLeuGI 293
882 .....GTCAACCAATTCGCCCGCAGTAACTTCGCTATCTTCGACCCAA 838
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 lalysValThrIleLeuAlaArgSerThrLeu...PhePheArgGlnuSp 308
837 AACGTCACGCAAGAGCG.....CGGTTGTGA 809
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 proAlaIleGlyGlnAlaValAlaThrAlaAlaPheArgMetGlnGlyIleGI 325
808 CTTGACAGAACCCCTAGGCAATCAGCGCTCGTGTTCAGACGCGCTGT 759
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 uValArg.....GluHisThrGlnAlaSerGlnValAlaIleTyI 338
758 GCAACAAACAGCCCAATGTAATTAATCTGATATATGATGTCACAC 709
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 leAsnGlyGlnGlyAspGlyGlnPheValLeuThr.....AlaHis 352
708 GGTTTATTCGCGCGCGCGCTGTGATGAATGATGTCGTCACATCA 659
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 GlyIleuLeuArgAlaAspLysLeuLeuValAlaThrGlyAlaAlaProAs 369
658 AACCGGACG.....ATCGCGGCGCGCG 636
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 nThrArgLysLeuAlaLeuAspAlaThrGlyValThrIleThrProGln 386
635 AATTCATGTTGTTGATGTTGCG.....AGCATTTTCAGACGCGCAC 595
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 lYAlaIleValIleAspProGlyMetArgThrSerValGlnHisIleTy 402
594 GTCTGCGCGCAGCTCTTACAAACATGATTTGCTGCTGGTCAACGCG 545
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 AlaIleGlyAspCysThrAspGlnProGlnPheValTy..... 415
544 TCATATCCACAGCGCGGTTGAAATCTCGCGCGCTTGTGATATG 495
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 .....ValAlaAla..AlaIleGlyThrArgAlaAlaIleAsnMet 428
494 ACCGTAGGCGTGGCAGCGCGGATGCTGC..... 462
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 ThrGlyGlyAspAlaAlaLeuAsnLeuThrAlaMetProAlaValAlaPh 445
461 .....ATCGCATTTGACGAAAGATGCGCAACGCGTCGCAATCGA 425
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
445 eThrAspProGlnValAlaThrValGlyTySerGlnAlaGlnAlaHisH 462
424 CGCAGCAATTTGCTGTAACGAGCGGTCGCGCAGCGCATGCCAACCG 375
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 lAspGlyIleLeuThrAspSerArgThrIleThrLeuAspAsValPro 478
374 GATTGATCAGGTTCCGCGCGCACTTCTCGCGGTTTAAGTTGCGACGCG 325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 ArgAlaLeuAlaAsnPheAspThrArgGlyPheIleLeuValValGI 495
324 TTCAGGTGCGGTCGTTCAACTCGATTTCGTCGTTCAACGCGCAA 275
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 uGlnGlySerGlyArgLeuIleGlyValGln..... 505
274 TCACGACTGATGAATACGCGCTTTTCGCCACGCTGAATGCGCGGAT 225
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 .....AlaVal 507
224 TTCCTGAGCGCGCGAGTAACACACGCGCGGATTTCTTTGTCTTC 175
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 AlaProGlnAlaGlyGlnLeuIleGlnThrAlaAlaLeuAlaIleArgAs 524

```

174 AACACGACTGGCTTTTGGAGCATCG..... CTTTCCTGACTT 131
 524 nargmetthrValGIngluLeuAlaAspGlnLeuPheProTyrLeuThrM 541
 130 TCATCGAGGG 120
 541 etValGInglu 544

seq_name: /cgn2_6/plodata/1/aa/5A_COMB.pep:US-08-427-097-20

seq_documentation_block:

Sequence 20, Application US/08427097

Patent No. 5668294

GENERAL INFORMATION:

APPLICANT: Meagher, Richard B.

APPLICANT: Sommers, Anne O.

TITLE OF INVENTION: Metal Resistance Sequences and

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSER: Greenlee, Winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder

STATE: CO

COUNTRY: US

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/427,097

FILING DATE: 21-APR-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Feiber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 40-94

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 564 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-427-097-20

alignment_scores:

Quality: 92.00 Length: 421
 Ratio: 0.479 Gaps: 17
 Percent Similarity: 45.606 Percent Identity: 20.190

alignment_block:

US-09-303-518D-125/rev x US-08-427-097-20 ..

Align seg 1/1 to: US-08-427-097-20 from: 1 to: 564

1122 GCGCGGTCGCGCGGTCGAGCGCTGTGAAGTGAAGTGTGTTT 1073
 163 GlyGlyLeuAlaAlaThrThrProThrLeuAlaArgThrAlaLeuAla 179
 1072 TCAGGAATGGCC.....GAGGCTGTAGCGCGTATGAG 1038
 179 aGIngluAlaArgValAspGlnLeuArgHisAlaLysTyrGIngluYI 196
 1037 TATTGTGCGGTCGCGCGCAACCGCAACGAGCTCTTT.....GCT 994
 196 Le.....LeuGIngluYAsnProAlaLeuThrValLeuHisGlySer 209

993 GCGGCTTCTTCGATACGGAATGTGATGTGTA.....GC 956
 210 AlaArgPheLysAspAsnArgAsnLeuLeuValGlnLeuAsnAspGlyGI 226
 955 GTCCCAATATATCGTCGCGCTGTGTATCGGCC..... 919
 226 yGIuArgValAlaAlaPheAspArgCysLeuIleAlaThrGlyAlaSerP 243
 918GTTCAATACCGCAACCGCA 901
 243 roAlaValProProIleProGlyLeuLysAspThrProTyrThrPheSer 259
 900 AATCAGCGCGTGTCTGT..... 883
 260 ThrGluAlaLeuValSerGluThrIleProLysArgLeuAlaValIleGI 276
 883 883
 276 ySerSerValAlaLeuGlnLeuAlaGlnAlaPheAlaArgLeuGIYA 293
 882GTCAACCAATTCGCGCGAGTAATTCGATCTTCGACCCAA 838
 293 IaLysValThrIleLeuAlaArgSerThrLeu...PhePheArgGluAsp 308
 837 AACGCTACGCAAGAGCG.....CGATTGTGA 809
 309 ProAlaIleGlyGlnAlaValThrAlaAlaPheArgMetGluGlyIleGI 325
 808 CTTCAGAACACCTAGGCGCAATACCGCGTGTGTTCAGAGCGCTGTT 759
 325 uValArg.....GluHisThrGlnAlaSerGlnValAlaTyrI 338
 758 GCAACAAACGCGCAATGTATATCATCTGATATGATGATGTCACAC 709
 338 LeasnGlyGlnGlyAspGlyGluPheValLeuThr.....AlaHis 352
 708 GGTATTATTCGCGCGCGCGCTCGATGAATGAATGCGTCCACTCA 659
 353 GlyGlnLeuArgAlaAspLysLeuValAlaThrGlyAlaGlyAlaProAs 369
 658 AACCGGCGAG.....ATGCGGCGCGCGG 636
 369 nThrArgLysLeuAlaLeuAspAlaThrGlyValThrLeuThrProGln 386
 635 AATCATGTGTTTCATGTGCG.....ACCATTTTCAGAGCGCAC 595
 386 lYAlaIleValIleAspProGlyMetArgThrSerValGlnHisIleTyr 402
 594 GTCTGCGCGAGCTGCTTACAAACATGATGATTTCGTCGTCGTCGTC 545
 403 AlaAlaGlyAspCysThrAspGlnProGlnPheValTyr..... 415
 544 TCAATACCAACAGCGCGCTTGAATCTCGCGGCGCTTCTTGAATAG 495
 416ValAlaAla..AlaAlaGlyThrArgAlaAlaIleAsnMet 428
 494 ACCGAGGTCGCGCGAGCGGATGTGTC..... 462
 429 ThrGlyLysAspAlaAlaLeuAsnLeuThrAlaMetProAlaValAlaPh 445
 461ATGCAATTGACGAAGATGCGCAACGCGTCCGCGCATCA 425
 445 eThrAspProGlnValAlaThrValGlyTyrSerGlnAlaGlnAlaHisH 462
 424 CGGCGAATTTTGTCTGAGAGGAGCGGTCGCGCAACGCAACGCG 375
 462 IsAspGlyIleLysThrAspSerArgThrLeuThrLeuAspAsnValPro 478
 374 GATTGATTCAGTTCGCGCGCACTTTCGCGCGCTTAAATGTTCCAGGCG 325
 479 ArgAlaLeuAlaAsnPheAspThrArgGlyPheIleLysLeuValValGI 495

```
324 TTCAGTCCGTAGCGTCAAACTGATTTGCTGCTTCAACGGCAA 275
      ::|||::: |||
495 uGlucIySerGIyArGIeuIlleGIyAlGln..... 505
274 TCACGACTGACTGAAGTACGGCGCTTTTCGCCACGGTGAATCGGGCGATT 225
      |||:::
506 .....AlaVal 507
224 TTGCCTGAAGCGCGCGCAGTAACACACACGCCGGATTCTTTTGTCTTC 175
      ||| ||| ::| ::| ::| ::| ::| ::|
508 AlaProGIuAlaGIyGIuLeuIlleGIuThrAlaAlaLeuAlaIleArgAs 524
174 AAACAGCACTGGCGCTTTTGTACGGCAATCG.....CCTTCCTTGACTT 131
      ::| ::| |||::: ||| |||:::
524 nArgmethIrrValGIuGIuLeuAlaAspGIuLeuPheProtyrLeuThrm 541
130 TCATCGAGGGG 120
      ::|::|::|
541 etValGIuGIy 544
```